

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: November 1, 2002, 12:14:41 ; Search time 26.5 seconds

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

**Title:** US-09-529-691A-1  
**Perfect score:** 92  
**Sequence:** I GVKGDKGNPGWPGAP 15  
**Scoring table:** BLOSUM62  
**Gapext:** 0.5

**Searched:** 747574 seqs, 11107196 residues  
**Total number of hits satisfying chosen parameters:** 747574

**Minimum DB seq length:** 0  
**Maximum DB seq length:** 2000000000

**Post-processing:** Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

**Database :**  
A\_Geneseq\_032802:\*

1: /SIDS1/gcdata/geneseq/geneseqp-emb1/AA1981.DAT:\*

2: /SIDS1/gcdata/geneseq/geneseqp-emb1/AA1982.DAT:\*

3: /SIDS1/gcdata/geneseq/geneseqp-emb1/AA1983.DAT:\*

4: /SIDS1/gcdata/geneseq/geneseqp-emb1/AA1984.DAT:\*

5: /SIDS1/gcdata/geneseq/geneseqp-emb1/AA1985.DAT:\*

6: /SIDS1/gcdata/geneseq/geneseqp-emb1/AA1986.DAT:\*

7: /SIDS1/gcdata/geneseq/geneseqp-emb1/AA1987.DAT:\*

8: /SIDS1/gcdata/geneseq/geneseqp-emb1/AA1988.DAT:\*

9: /SIDS1/gcdata/geneseq/geneseqp-emb1/AA1990.DAT:\*

10: /SIDS1/gcdata/geneseq/geneseqp-emb1/AA1991.DAT:\*

11: /SIDS1/gcdata/geneseq/geneseqp-emb1/AA1992.DAT:\*

12: /SIDS1/gcdata/geneseq/geneseqp-emb1/AA1993.DAT:\*

13: /SIDS1/gcdata/geneseq/geneseqp-emb1/AA1994.DAT:\*

14: /SIDS1/gcdata/geneseq/geneseqp-emb1/AA1995.DAT:\*

15: /SIDS1/gcdata/geneseq/geneseqp-emb1/AA1996.DAT:\*

16: /SIDS1/gcdata/geneseq/geneseqp-emb1/AA1997.DAT:\*

17: /SIDS1/gcdata/geneseq/geneseqp-emb1/AA1998.DAT:\*

18: /SIDS1/gcdata/geneseq/geneseqp-emb1/AA1999.DAT:\*

19: /SIDS1/gcdata/geneseq/geneseqp-emb1/AA2000.DAT:\*

20: /SIDS1/gcdata/geneseq/geneseqp-emb1/AA2001.DAT:\*

22: /SIDS1/gcdata/geneseq/geneseqp-emb1/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

**Result No.** **Score** **Query Match Length** **DB ID** **Description**

1	92	100.0	15	12	AAR12973	RESULT 1 ID AAR12973 XX standard; peptide: 15 AA. AC AAR12973; XX DT 24-SEP-1991 (first entry) DE IV-H1 based on type IV collagen alpha 1 chain triple helical region. XX KW malignant cell growth; melanoma cell motility; cellular adhesion. XX OS Synthetic. XX PN W09108755-A. XX PD 27-JUN-1991. XX PF 06-DEC-1990; 90WO-US07162. XX PR 14-DEC-1989; 89US-0450419.
2	92	100.0	15	16	AAR83639	XX (MINU ) MINNESOTA UNIVERSITY.
3	92	100.0	15	16	AAR82922	XX Chelberg MK, Tsilibary PE, McCarthy JB;
4	92	100.0	15	17	AAR01139	XX DR WPI; 1991-207864/28.
5	92	100.0	15	18	AAR13022	XX Peptidopeptide for promoting tumour cell adhesion, spreading and
6	92	100.0	15	19	AAR56114	XX Peptidopeptide for promoting tumour cell adhesion, spreading and
7	92	100.0	15	20	AYV04481	XX Peptidopeptide for promoting tumour cell adhesion, spreading and
8	92	100.0	15	21	AYV78574	XX Peptidopeptide for promoting tumour cell adhesion, spreading and
9	92	100.0	24	19	AAR46044	XX Peptidopeptide for promoting tumour cell adhesion, spreading and
10	92	100.0	27	18	AAR13025	XX Peptidopeptide for promoting tumour cell adhesion, spreading and
11	92	100.0	39	19	AAW46046	XX Peptidopeptide for promoting tumour cell adhesion, spreading and

CC This polypeptide (designated IV-H1) corresponds to type IV collagen residues 1263-1277 from the major triple helical region of the alpha 1 chain type of IV collagen. It, or a polypeptide comprising this sequence, can be used to promote cellular attachment to substrata or to inhibit the metastasis and invasion of tumour cells.

CC SQ Sequence 15 AA;

Query Match 100.0%; Score 92; DB 12; Length 15;  
Best Local Similarity 100.0%; Pred. No. 5.2e-07;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 1 GVKGDKGNPGWPGAP 15

RESULT 2  
AAR83639  
ID AAR83639 standard; peptide; 15 AA.  
XX  
AC AAR83639;  
XX  
DT 11-APR-1996 (first entry)

DE Collagen IV alpha-2 chain NCL domain peptide IV-H1 (1263-1277).  
XX  
KW Collagen; type IV; alpha-2 chain; NCL domain; fibroblast; cell; proliferation; suppression; epithelial; glial; epiretinal; eye; membranes; contraction; migration; vitreoretinopathy; scarring; KW  
vitreoretinal; glaucoma.  
XX  
OS Synthetic.  
XX  
WO522979-A1.

PN  
XX  
PD 04-APR-1995.

PF 17-SEP-1993; 93JP-0254779.

PR 17-SEP-1993; 93JP-0254779.

XX  
PA (JAPG ) NIPPON ZEON KK.  
XX  
DR WPI; 1995-167254/22.

XX  
PT Cancer metastasis inhibitive peptide derivs. - useful for inhibition of cancer metastasis, healing of wounds and regulation of PT immunogenicity.  
XX  
PS Disclosure; Page 3; 6pp; Japanese.

XX  
CC The peptides AAR70472-90 and AAR62907-24 are peptide derivatives which inhibit cancer metastasis. They are composed of an adhesive peptide with a core sequence selected from: RGD (AAR70472-85), YIGSR (AAR70486-90) or other sequence (AAR83907-24), linked to a water soluble polysaccharide, preferably a water soluble dextran, at the C-terminus. CC  
The peptides are useful in inhibiting cancer metastasis, healing wounds CC  
and the regulation of immunogenicity.  
XX  
PA (MINU ) UNIV MINNESOTA.  
XX  
PT Agarwal A, Balles M, Furchtgott LT, Gregerson DS, Murali S;  
PI Wright MM, Skubitz AP;  
XX  
DR WPI; 1995-320291/41.

XX  
PT Treating proliferative vitreoretinopathy, retinal scarring and glaucoma - using peptide derived from fibronectin, laminin or collagen  
XX  
PS Example 1; Page 29; 84bp; English.

XX  
CC A polypeptide which includes at least five amino acids from a fragment of the fibronectin A chain (AAR83679-84), C-terminal G domain of the laminin A chain (AAR83642-78) or NCL domain of the alpha-2 chain of type IV collagen (AAR83639-41) suppresses fibroblast, epithelial and glial cell proliferation, contraction and migration within the eye. It may be used in the treatment of proliferative vitreoretinopathy, vitreoretinal scarring and glaucoma.  
XX  
SQ Sequence 15 AA;

Query Match 100.0%; Score 92; DB 16; Length 15;  
Best Local Similarity 100.0%; Pred. No. 5.2e-07;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 1 GVKGDKGNPGWPGAP 15

RESULT 3  
AAR82922  
ID AAR82922 standard; peptide; 15 AA.  
XX  
AC AAR82922;  
XX  
DT 20-DEC-1995 (first entry)

DE Non-RGD, non-YIGSR cancer metastasis inhibitory peptide #16.  
XX  
KW Cancer metastasis; adhesive peptide; core sequence; dextran; cancer; water soluble polysaccharide; metastasis; wound; immunogenicity.  
XX  
OS Synthetic.  
XX  
PN JP07089399-A.  
XX  
PD 04-APR-1995.

PF 17-SEP-1993; 93JP-0254779.

PR 17-SEP-1993; 93JP-0254779.

XX  
PA (JAPG ) NIPPON ZEON KK.  
XX  
DR WPI; 1995-167254/22.

XX  
PT Cancer metastasis inhibitive peptide derivs. - useful for inhibition of cancer metastasis, healing of wounds and regulation of PT immunogenicity.  
XX  
PS Disclosure; Page 3; 6pp; Japanese.

XX  
CC The peptides AAR70472-90 and AAR62907-24 are peptide derivatives which inhibit cancer metastasis. They are composed of an adhesive peptide with a core sequence selected from: RGD (AAR70472-85), YIGSR (AAR70486-90) or other sequence (AAR83907-24), linked to a water soluble polysaccharide, preferably a water soluble dextran, at the C-terminus. CC  
The peptides are useful in inhibiting cancer metastasis, healing wounds CC  
and the regulation of immunogenicity.  
XX  
PA (MINU ) UNIV MINNESOTA.  
XX  
PT Agarwal A, Balles M, Furchtgott LT, Gregerson DS, Murali S;  
PI Wright MM, Skubitz AP;  
XX  
DR WPI; 1995-320291/41.

XX  
PT Treating proliferative vitreoretinopathy, retinal scarring and glaucoma - using peptide derived from fibronectin, laminin or collagen  
XX  
PS Example 1; Page 29; 84bp; English.

XX  
CC A polypeptide which includes at least five amino acids from a fragment of the fibronectin A chain (AAR83679-84), C-terminal G domain of the laminin A chain (AAR83642-78) or NCL domain of the alpha-2 chain of type IV collagen (AAR83639-41) suppresses fibroblast, epithelial and glial cell proliferation, contraction and migration within the eye. It may be used in the treatment of proliferative vitreoretinopathy, vitreoretinal scarring and glaucoma.  
XX  
SQ Sequence 15 AA;

Query Match 100.0%; Score 92; DB 16; Length 15;  
Best Local Similarity 100.0%; Pred. No. 5.2e-07;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Db 1 GVKGDKGNPGWPGAP 15

RESULT 4  
AAW01139  
ID AAW01139 standard; peptide; 15 AA.  
XX  
AC AAW01139;  
XX  
DT 18-DEC-1996 (first entry)  
XX  
DE Peptide 10 for glia cell removal derived from collagen.  
XX  
KW glia cell; neuron; analysis; behaviour; selective; removal.  
KW laminin; fibronectin; collagen.  
XX  
OS Synthetic.  
XX  
FH Key Modified-site 15 Location/Qualifiers  
FT /note= "Pro-NH2"

XX JP08073495-A.  
 PN XX  
 XX 19-MAR-1996.  
 PD XX  
 PR XX 31-AUG-1994; 94JP-0232263.  
 PT XX 31-AUG-1994; 94JP-0232263.  
 XX PA (AGEN ) AGENCY OF IND SCI & TECHNOLOGY.  
 XX DR WPI; 1996-205531/21.  
 XX PT Selective remover for glia cells from neuronal cell cultures -  
 CC useful for studying behaviour of neurons in the absence of glia  
 PT cells.  
 XX Example 1; Page 4; 6pp; Japanese.  
 PS XX AAW01135-40 are peptides, derived from laminin, fibronectin and  
 CC collagen, which selectively remove glia cells in the presence of neurons.  
 CC This enables analysis of behaviour of neurons in the absence of glia  
 CC cells.

SQ Sequence 15 AA:  
 Query Match 100.0%; score 92; DB 17; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 5.2e-07; Mismatches 0; Indels 0; Gaps 0;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 CC OY 1 GVKGDKGNPGWGP 15  
 Db 1 GVKGDKGNPGWGP 15

RESULT 5  
 AAW13022  
 ID AAW13022 standard; peptide; 15 AA.  
 XX AC AAW13022;  
 XX DT 30-APR-1997 (first entry)  
 XX DE Collagen alpha1(IV) chain residues 1263-1277.  
 XX KW Collagen; cell; adhesion; migration; promotion; alpha1(IV) chain;  
 XX triple helical polypeptide; solid phase synthesis; study;  
 XX O-glycosidic bond; deuterium label; structure; biological activity.  
 XX OS Homo sapiens.  
 XX OS US5576419-A.  
 XX PD 19-NOV-1996.  
 XX PP 30-JUN-1993; 93US-0085633.  
 XX PR 30-JUN-1993; 93US-0085633.  
 XX PR 27-SEP-1995; 95US-0534342.  
 XX PA (MINU ) UNTV MINNESOTA.  
 XX PI Fields GB;  
 XX DR WPI; 1997-011309/01.

PT Solid Phase synthesis of triple-helical branched polypeptide -  
 PT under O-glycosidic bond and deuterium label retaining conditions,  
 PT which may contain collagen cell adhesion sequences, useful for  
 PT studying structure and biological activity of collagen  
 XX PS Claim 18; Column 25; 25pp; English.

CC The present peptide is a collagen cell adhesion sequence (CCAS),  
 CC known to promote adhesion and migration of various cells,  
 CC comprising residues 1263-1277 of the collagen alpha1(IV) chain.  
 CC The triple helical polypeptide 2 (THP-2), which includes the  
 CC present CCAS (residues 25-39) in its triple helix inducing region,  
 CC was prepared by solid phase synthesis under O-glycosidic bond and  
 CC deuterium label retaining conditions. THP-2 can be used to study  
 CC the structure and biological activity of collagen.

SQ Sequence 15 AA:  
 Query Match 100.0%; Score 92; DB 18; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 5.2e-07; Mismatches 0; Indels 0; Gaps 0;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 CC OY 1 GVKGDKGNPGWGP 15  
 Db 1 GVKGDKGNPGWGP 15

RESULT 6  
 AAW56114  
 ID AAW56114 standard; peptide; 15 AA.  
 XX AC AAW56114;  
 XX DT 09-JUL-1998 (first entry)  
 XX DE Peptide representing the alpha-1(IV)1263-1277 collagen sequence.  
 XX KW Alpha-1(IV)1263-1277 collagen sequence; peptide-amphiphile complex;  
 KW lipophilic portion; stable lipid film; bilayer membrane system;  
 KW drug-targeting; melanoma cell.  
 XX OS Synthetic.  
 XX PN WO9807752-A1.  
 XX PD 26-FEB-1998.  
 XX PF 27-AUG-1996; 96WO-US14145.  
 XX PR 23-AUG-1996; 96US-0702254.  
 XX PA (MINU ) UNTV MINNESOTA.  
 XX PI Fields GB, Tirrell MW;  
 XX DR WPI; 1998-179051/16.  
 XX PT Peptide-amphiphile complexes which are able to form micelles or  
 PT vesicles - comprise a lipophilic portion and a peptide portion which  
 PT has a secondary structure  
 XX PS Claim 15; Page 23; 33pp; English.  
 XX The present sequence represents the alpha-1(IV)1263-1277 collagen  
 CC sequence. This sequence is used to construct a peptide-amphiphile  
 CC complex to exemplify the invention. The peptide-amphiphile complex of  
 CC the specification comprises a lipophilic portion and a peptide portion  
 CC having a secondary structure. The peptide portion preferably has  
 CC biological activity, e.g. cell recognition activity or enzymatic  
 CC activity. The complex may be used for forming stable lipid films on  
 CC substrates, or may be used in bilayer membrane systems. It may, in  
 CC micellar or vesicular form, be used in preparation of drug-targeting  
 CC systems against melanoma cells.

SQ Sequence 15 AA:  
 Query Match 100.0%; score 92; DB 19; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 5.2e-07; Mismatches 0; Indels 0; Gaps 0;







AAW13024  
 AAW13024 standard; peptide; 43 AA.  
 ID  
 CC  
 AC  
 XX  
 AAW13024;  
 DT  
 30-APR-1997 (first entry)  
 XX  
 DE Collagen alpha1(IV) chain containing triple helical polypeptide 2.  
 XX  
 KW Triple helical polypeptide; collagen; cell; adhesion; migration;  
 KW promotion; alpha1(IV) chain; solid phase synthesis; study;  
 KW o-glycosidic bond; deuterium label; structure; biological activity;  
 KW homotrimeric; heterotrimeric; nuclear magnetic resonance; NMR.  
 XX  
 OS synthetic.  
 XX  
 FH Key Location/Qualifiers  
 FT Modified-site 3  
 FT /label= HYP  
 FT Modified-site 6  
 FT /label= HYP  
 FT Modified-site 9  
 FT /label= HYP  
 FT Modified-site 12  
 FT /label= HYP  
 FT Modified-site 15  
 FT /label= HYP  
 FT Modified-site 18  
 FT /label= HYP  
 FT Modified-site 21  
 FT /label= HYP  
 FT Modified-site 24  
 FT /label= HYP  
 FT Peptide 25..39  
 FT /note= "collagen alpha1(IV) chain residues  
 FT 1263-1277"  
 FT Modified-site 40  
 FT /note= "linked via peptide bond to Pro39 of peptide  
 sequence"  
 FT Modified-site 41  
 FT /note= "linked via peptide bond to Pro39 of peptide  
 comprising residues 1-39 of present  
 sequence"  
 FT US5576419-A.  
 XX PD 19-NOV-1996.  
 XX PF 30-JUN-1993;  
 XX PR 30-JUN-1993;  
 XX PR 27-SEP-1995;  
 XX PA (MINU ) UNIV MINNESOTA.  
 XX PI Fields GB;  
 XX DR WPI; 1997-011309/01.  
 XX PT Solid phase synthesis of triple-helical branched polypeptide -  
 under O-glycosidic bond and deuterium label retaining conditions,  
 which may contain collagen cell adhesion sequences, useful for  
 studying structure and biological activity of collagen  
 XX PS Claim 48; column 27; 25pp; English.  
 XX  
 CC The present sequence is that of triple helical polypeptide 2  
 (THP-2), which includes a collagen cell adhesion sequence, known to  
 promote adhesion and migration of various cells (residues 1263-1277  
 of the collagen alpha1(IV) chain), in its triple helix inducing  
 region. THP-2 was prepared by solid phase synthesis under  
 o-glycosidic bond and deuterium label retaining conditions, without  
 overlapping NMR resonances.  
 XX SQ Sequence 43 AA:  
 Query Match 100.0%; Score 92; DB 18; Length 43;  
 Best Local Similarity 100.0%; Prd. No. 1.6e-06;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 25 GVKGDKGNPGWPgap 39  
 XX RESULT 14  
 AAM40863  
 ID AAM40863 standard; Protein; 1669 AA.  
 XX AC AAM40863;  
 XX DT 22-OCT-2001 (first entry)  
 XX DE Human polypeptide SEQ ID NO 5794.  
 XX KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;  
 KW peripheral nervous system; neuropathy; central nervous system; CNS;  
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
 KW leukaemia.  
 XX OS Homo sapiens.  
 XX PN WO200153312-A1.  
 XX PD 26-JUL-2001.  
 XX PF 26-DEC-2000; 2000WO-US34263.  
 XX PR 21-JAN-2000; 2000US-0488725.  
 XX PR 25-APR-2000; 2000US-052317.  
 XX PR 09-JUL-2000; 2000US-0508042.  
 XX PR 19-JUL-2000; 2000US-0620312.  
 XX PR 03-AUG-2000; 2000US-053450.  
 XX PR 14-SEP-2000; 2000US-062191.  
 XX PR 19-OCT-2000; 2000US-0693036.  
 XX PR 29-NOV-2000; 2000US-0727344.  
 XX PA (HYSE-) HYSEQ INC.  
 XX PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;  
 XX DR WPI; 2001-442253/47.  
 XX DR N-FSDB; AAI60019.  
 XX  
 CC The invention relates to human nucleic acids (AA157798-AA161369) and  
 the encoded polypeptides (AAM38642-AAM42213) with nootropic,  
 immunosuppressant and cytostatic activity. The polynucleotides are useful  
 in gene therapy. A composition containing a polypeptide or polynucleotide  
 of the invention may be used to treat diseases of the peripheral nervous  
 system, such as peripheral nervous injuries, peripheral neuropathy and  
 localised neuropathies and central nervous system diseases, such as

CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager Syndrome. Other uses include the utilisation of the activities such as: immune system suppression, CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic and thrombolytic activity, cancer diagnosis and therapy, drug screening, assays for receptor activity, arthritis and inflammation, leukaemias and C.N.S disorders.

Note: The sequence data for this patent did not form part of the printed specification.

SQ sequence 1669 AA;

Query Match 100.0%; Score 92; DB 22; Length 1669;  
Best Local Similarity 100.0%; Pred. No. 7.5e-05;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVKGDKGNPGWPGAP 15  
Db 1263 GVKGDKGNPGWPGAP 1277

RESULT 15

AM39077 standard; Protein; 1672 AA.  
ID AM39077;  
XX DT 22-OCT-2001 (first entry)

DE Human polypeptide SEQ ID NO 2222.  
XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer; peripheral nervous system; neuropathy; central nervous system; CNS; Alzheimer's; Parkinson's disease; Huntington's disease; hemostatic; amyotrophic lateral sclerosis; Shy-Drager syndrome; chemotactic; chemokinetic; thrombolytic; drug screening; arthritis; inflammation; leukæmia.

KW XX Homo sapiens.  
OS XX WO200153312-A1.  
PN XX PD 26-JUL-2001.  
XX PR 26-DEC-2000; 2000WO-US34263.

PR 21-JAN-2000; 2000US-0488725.  
PR 25-APR-2000; 2000US-055317.  
PR 09-JUL-2000; 2000US-059042.  
PR 19-JUL-2000; 2000US-0620312.  
PR 03-AUG-2000; 2000US-063450.  
PR 14-SEP-2000; 2000US-0662191.  
PR 19-OCT-2000; 2000US-0693036.  
PR 29-NOV-2000; 2000US-0727344.

XX PA (HYSE-) BYSEQ INC.  
XX PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Oian XB, Ren F, Wang D; PT Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J; XX Zhao QA, Zhou P, Goodrich R, Dimancic RT; DR WPI; 2001-442253/47. N-PSDB; AA15233.

XX PT Novel nucleic acids and polypeptides, useful for treating disorders such as central nervous system injuries - Example 4; SEQ ID NO 2222; 10078pp; English.

XX The invention relates to human nucleic acids (AA157798-AA161369) and the encoded polypeptides (AAm3862-AAm42213) with nootropic, immunosuppressant and cytostatic activity. The polynucleotides are useful in gene therapy. A composition containing a polypeptide or polynucleotide

CC of the invention may be used to treat diseases of the peripheral nervous system, such as peripheral nervous injuries, peripheral neuropathy and localised neuropathies and central nervous system diseases, such as Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager Syndrome. Other uses include the utilisation of the activities such as: immune system suppression, Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic and thrombolytic activity, cancer diagnosis and therapy, drug screening, assays for receptor activity, arthritis and inflammation, leukaemias and C.N.S disorders.

Note: The sequence data for this patent did not form part of the printed specification.

XX Sequence 1672 AA;  
SQ

Query Match 100.0%; Score 92; DB 22; Length 1672;  
Best Local Similarity 100.0%; Pred. No. 7.5e-05;  
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GVKGDKGNPGWPGAP 15  
Db 1266 GVKGDKGNPGWPGAP 1280

RESULT 16

AAB50169  
ID AAB8169 standard; Protein; 406 AA.  
XX AC AAB8169;  
XX DT 14-MAR-2001 (first entry)

DE Lung cancer associated polypeptide sequence SEQ ID 507.  
XX Human; lung cancer associated protein; neuroprotective; cytostatic; cardioactive; immunomodulatory; muscular active; vulnerable; gastrointestinal; nephrotropic; antiinfective; gynecological; antibacterial; diagnosis; neural disorder; immune disorder; reproductive; proliferative disorder; wound healing; infectious disease.

KW XX Homo sapiens.  
OS XX WO200055180-A2.  
PN XX PD 21-SEP-2000.

PR 08-MAR-2000; 2000WO-US05918.  
PR 12-MAR-1999; 9905-0124270.  
PR (HUWA-) HUMAN GENOME SCI INC.  
PA (ROSE/) ROSEN C A.  
XX PI Ruben SM;  
XX DR WPI; 2000-587514/55.  
DR N-PSDB; AA18045.

XX PT Lung cancer associated gene sequences, referred to as lung cancer antigens, useful for treatment, prevention, and diagnosis of disorders such as lung cancer -

XX PS Claim 11; Page 996-998; 1425pp; English.

XX CC Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer associated proteins represented in AAB58106 - AAB5848. Lung cancer associated proteins and polynucleotide sequences, their agonists, and antagonists may have neuroprotective; cytotoxic; cardioactive; immunomodulatory; muscular active general; pulmonary; gastrointestinal; general; nephrotoxic; antiinfective; gynaecological; or antibacterial activity. The invention also includes antibodies specific for the protein or polynucleotide sequences. The lung cancer associated polynucleotide sequences may be used for detection of lung cancer,

chromosome identification, as chromosome markers, and for numerous other diagnostic or research purposes. The proteins may be used to treat disorders such as neural, immune, muscular, reproductive, gastrointestinal, pulmonary, cardiovascular, renal, and proliferative disorders. The proteins may also be used in the treatment of wounds and infectious diseases. Polynucleotide sequences AAF1425 - AAF1843 and peptide AAB58549 are used in the course of the invention for the identification and characterisation of the polynucleotide and protein sequences.

Sequence 406 AA;

Query Match 93.5%; Score 86; DB 21; Length 406;  
Best Local Similarity 100.0%; Pred. No. 0.00013; Mismatches 0; Indels 0; Gaps 0;

QY 2 VKGDKGNGPGWPGAP 15  
Db 1 VKGDKGNGPGWPGAP 14

RESULT 17

ID AAR23873 standard; Protein: 772 AA.  
XX AAR23873;  
XX DT 25-NOV-1992 (first entry)  
DE Human alpha 5 (IV) of type IV collagen.  
XX Mutations: Alport's syndrome; basement membranes; diabetes mellitus.  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT Misc-difference 43..47  
FT /note= "interruption in Gly-X-Y sequence"  
FT Misc-difference 159..160  
FT /note= "interruption in Gly-X-Y sequence"  
FT Misc-difference 275..277  
FT /note= "interruption in Gly-X-Y sequence"  
FT Misc-difference 334..336  
FT /note= "interruption in Gly-X-Y sequence"  
FT Misc-difference 456..458  
FT /note= "interruption in Gly-X-Y sequence"  
PN US511440-A.  
XX 19-MAY-1992.  
XX PD 07-JUL-1989; 89US-0377238.  
PR 07-JUL-1989; 89US-0377238.  
XX (TRYG/) TRYGGVASON K.  
PI Hostikka S L, Tryggvason K;  
XX WPI: 1992-192174/23.  
DR N-PSDB; AAQ24551.  
XX Isolation of DNA encoding alpha-5(IV)polypeptide of type IV collagen - to detect mutations in genes for alpha-5(IV) chain which produce genetic or acquired basement membrane disorders e.g. Alport's syndrome

Disclosure; Fig 2; 14pp; English.

XX The sequence is that of the alpha 5(IV) polypeptide chain of human type IV collagen, the major component of basement membranes. The protein contains the Gly-X-Y repeat coding sequence typical for collagenous proteins at one end and a typical NC-domain coding

CC sequence at the other end. The sequence can be used to detect CC mutations in individual genes specific for this chain which can, CC directly or indirectly, produce several human diseases. It can CC also be used to determine genetic, e.g. Alport's syndrome, or CC acquired e.g. diabetes mellitus, disorders of the basement membrane, CC and as probes or antibodies against these nucleotide sequences. CC Gene fragments generated through amplifications from human genomic CC or cloned DNA can also be used for detection and analysis of genes. XX

Sequence 772 AA;

Query Match 75.0%; Score 69; DB 13; Length 772;  
Best Local Similarity 73.3%; Pred. No. 0.074; Mismatches 2; Indels 0; Gaps 0;

QY 1 GVKGDKGNPGWPGAP 15  
Db 366 GVKGEKGNPGGCGLP 380

RESULT 18

ID AAW09643 standard; Protein: 772 AA.  
XX AAW09643;  
XX DT 16-JUN-1997 (first entry)  
DE Human type IV collagen alpha-5.  
XX Collagen alpha5(IV); basement membrane; Alport's syndrome; KW nephritis; kidney; renal failure; antibody; diagnosis; COL4A5 gene; X chromosome.  
XX OS Homo sapiens.  
FH Key Location/Qualifiers  
FT Domain 1..543  
FT /label= Collagenous domain contains Gly-X-Y  
FT /note= "collagenous domain contains Gly-X-Y tripeptide repeat; interrupted at positions 43-47, 159-160, 275-276, 334-335, 456-459"  
FT Domain 544..772  
FT /label= Non-collagenous\_domain 742..751  
FT Peptide /label= Immunogenic\_peptide  
FT /note= "peptide used to raise diagnostic antibodies (Claim 1)"  
FT  
XX PN US5593900-A.  
XX PD 14-JAN-1997.  
XX PF 07-JUL-1989; 89US-0377238.  
XX PR 20-DEC-1990; 90US-0530563.  
PR 07-JUL-1989; 89US-0377238.  
PR 11-OCT-1994; 94US-0321084.  
XX PA (HOST/ HOSTIKKA S L.  
PA (HOST/ HOHTYA M.  
PA (TRYG/) TRYGGVASON K.  
XX PI Hostikka S L, HOHTYA M, Tryggvason K;  
XX DR WPI: 1997-099481/09.  
DR N-PSDB; AAQ24551.  
XX PR New antibodies specific for human type IV collagen alpha5 chain -  
PT used to detect absence of this chain in patients with renal failure  
XX Disclosure; Fig 2A-2B; 12pp; English.

PS

Page 10

The amino acid sequence of a portion (AAW03643) of the previously unknown human type IV collagen chain, alpha5(IV), was deduced from cDNA clones (see also AAT47812) obtdd. using Probes based on conserved sequences of human alpha1(IV) and alpha2(IV) collagen chains and of the Drosophila alpha(IV) chain. It includes a complete non collagenous domain that shows 83% identity with that of alpha1(IV) and 63% with that of the alpha2(IV) chain. Mutation CC in the alpha5(IV) gene (COLA5) are associated with Alport's syndrome. Antibodies raised against a peptide (see also AAW03644) specific to alpha5(IV) can be used in the diagnosis of basement membrane disorders such as Alport's syndrome.

CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABC0010-ABC0377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at [ftp.wipo.int/pub/published-pct/sequences](http://ftp.wipo.int/pub/published-pct/sequences).

SQ	Sequence	772 AA;
Query Match		75.0%; Score 69; DB 18; Length 772;
Best Local Similarity		73.3%; Pred. No. 0.074;
Matches	11; Conservative	2; Mismatches 2; Indels 0; Gaps 0
QY	1 GVKDDKGNGPWPQAP 15  : :          Db 366 GIKEGKGNPGQGLP 380	

```

Query Match          75.0%; Score 69; DB 22; Length 1685;
Best Local Similarity 73.3%; Pred. No. 0.17;
Matches 11; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
Qy      1 GVKGDKGPNPGPAG 15
       ||||||| | |
Db      1279 GIKGEKGKPNPGPLP 1293

```

CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG0010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://wipo.int/pub/published_pct_sequences).

SQ Sequence 1693 AA;

Query Match 75.0%; Score 69; DB 22; Length 1693;

Best Local Similarity 73.3%; Pred. No. 0.17; Matches 11; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 GVKGDKGNPGNGPAGAP 15  
 | : : : : : | | | | |  
 Db 1287 GIKGEKGNPGCPGLP 1301

RESULT 21

AAB58180  
 ID AAB58180 standard; Protein: 430 AA.

AC AAB58180;  
 XX  
 DT 14-MAR-2001 (first entry)

XX Lung cancer associated polypeptide sequence SEQ ID 518.  
 DE Human: lung cancer associated protein; neuroprotective; cytostatic;  
 KW cardiotropic; immunomodulatory; muscular active; vulnerability;  
 KW gastrointestinal; nephrotoxic; antiinfective; gynecological;  
 KW antibacterial; diagnosis; neural disorder; immune disorder; reproductive;  
 KW proliferative disorder; wound healing; infectious disease.

OS Homo sapiens.  
 XX  
 PN WO20055180-A2.

PD 21-SEP-2000.

XX  
 PF 08-MAR-2000; 2000WO-US05918.

XX  
 PR 12-MAR-1999; 99US-0124270.

XX  
 PA (HUMA-) HUMAN GENOME SCI INC.

PA (ROSE/-) ROSEN C A.

XX  
 PT Ruben SM;

XX  
 DR WPI; 2000-587514/55.

DR N-PSDB; AAF18056.

XX  
 PS Disclosure; SEQ ID NO 33723; 21pp + Sequence Listing; English.

XX  
 PT Lung cancer associated gene sequences, referred to as lung cancer  
 PT antigens, useful for treatment, prevention, and diagnosis of disorders  
 PT such as lung cancer -

PS Claim 11; Page 108-1010; 1425pp; English.

XX  
 CC Polynucleotide sequences AAF17982 - AAF8424 encode human lung cancer  
 CC associated proteins represented in AAB58106 - AAB58548. Lung cancer  
 CC associated proteins and polynucleotide sequences, their agonists, and  
 CC antagonists may have neuroprotective; cytosolic; cardiotropic;  
 CC immunomodulatory; muscular active; general; pulmonary; gastrointestinal  
 CC general; nephrotoxic; antiinfective; gynecological; or antibacterial  
 CC activity. The invention also includes antibodies specific for the  
 CC protein or polynucleotide sequences. The lung cancer associated  
 CC polynucleotide sequences may be used for detection of lung cancer,  
 CC chromosome identification, as chromosome markers, and for numerous other  
 CC diagnostic or research purposes. The proteins may be used to treat  
 CC disorders such as neural, immune, muscular, reproductive,  
 CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative

CC disorders. The proteins may also be used in the treatment of wounds and  
 CC infectious diseases. Polynucleotide sequences AAF8425 - AAF18433 and  
 CC peptide AB5849 are used in the course of the invention for the  
 CC identification and characterisation of the polynucleotide and protein  
 CC sequences.

SQ Sequence 430 AA;

Query Match 69.6%; Score 64; DB 21; Length 430;

Best Local Similarity 73.3%; Pred. No. 0.21; Matches 11; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 GVKGDKGNPGNGPAGAP 15  
 | : : : : | | | | |  
 Db 86 GPKGPCKDPGPGPAG 100

RESULT 22

ABB68977  
 ID ABB68977 standard; Protein: 224 AA.

AC ABB68977;  
 XX  
 DT 26-MAR-2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID NO 33723.  
 DE Drosophila: developmental biology; cell signalling; insecticide;  
 KW pharmaceutical.

OS Drosophila melanogaster.  
 XX  
 PN WO200171042-A2.

PD 27-SEP-2001.

XX  
 PF 23-MAR-2001; 2001WO-US09231.

XX  
 PR 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX  
 PA (PEKE ) PE CORP NY.

XX  
 PI Venter JC, Adams M, Li PWD, Myers EW;

XX  
 DR WPI; 2001-656860/75.

DR N-PSDB; ABL13080.

XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -

XX  
 PS Disclosure; SEQ ID NO 33723; 21pp + Sequence Listing; English.

XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL01840-ABL16175) and the encoded proteins  
 CC sequences (ABL01840-ABL16175) and the encoded proteins  
 CC (ABB5773-ABB72072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://wipo.int/pub/published_pct_sequences).

XX  
 SQ Sequence 224 AA;

Query Match 66.3%; Score 61; DB 22; Length 224;

Best Local Similarity 66.7%; Pred. No. 0.29; Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 1 GVKGDKGNPGNGPAG 15

Db 118 | :||||:||| ||| |  
 RESULT 23  
 AAU16066 ID AAU16066 standard; protein; 429 AA.  
 XX AC AAU16066;  
 XX DT 07-NOV-2001 (first entry)  
 XX DE Human novel secreted protein, Seq ID 1019.  
 XX KW Human; immunosuppressive; antiarthritic; antirheumatic;  
 KW cytostatic; cardiotonic; vasotropic; cerebroprotective; nootropic;  
 KW neuroprotective; antibacterial; virucide; fungicide; ophthalmological;  
 KW vulnerability; secreted protein; rheumatoid arthritis;  
 KW hyperproliferative disorder; cardiovascular disorder; cardiac arrest;  
 KW cerebrovascular disorder; cerebral ischaemia; angiogenesis;  
 KW nervous system disorder; Alzheimer's disease; infection; ocular disorder;  
 KW corneal infection; wound healing; epithelial cell proliferation;  
 KW skin ageing; food additive; preservative; antiproliferative.  
 XX OS Homo sapiens.  
 XX PN WO20015322-A2.  
 XX PD 02-AUG-2001.  
 XX PR 17-JAN-2001; 2001WO-US01341.  
 XX PR 31-JAN-2000; 2000US-0179065.  
 PR 04-FEB-2000; 2000US-0180628.  
 PR 24-FEB-2000; 2000US-0184664.  
 PR 02-MAR-2000; 2000US-0180350.  
 PR 16-MAR-2000; 2000US-0189874.  
 PR 17-MAR-2000; 2000US-0190076.  
 PR 18-APR-2000; 2000US-198123.  
 PR 19-MAY-2000; 2000US-020515.  
 PR 07-JUN-2000; 2000US-0209467.  
 PR 28-JUN-2000; 2000US-0214886.  
 PR 30-JUN-2000; 2000US-0215135.  
 PR 07-JUL-2000; 2000US-0216647.  
 PR 11-JUL-2000; 2000US-0215880.  
 PR 11-JUL-2000; 2000US-0217487.  
 PR 14-JUL-2000; 2000US-0218290.  
 PR 26-JUL-2000; 2000US-0220963.  
 PR 14-AUG-2000; 2000US-0224154.  
 PR 14-AUG-2000; 2000US-0224519.  
 PR 14-AUG-2000; 2000US-0225213.  
 PR 14-AUG-2000; 2000US-0225214.  
 PR 14-AUG-2000; 2000US-0225266.  
 PR 14-AUG-2000; 2000US-0225267.  
 PR 14-AUG-2000; 2000US-0225269.  
 PR 22-AUG-2000; 2000US-0225447.  
 PR 14-AUG-2000; 2000US-0225757.  
 PR 14-AUG-2000; 2000US-022718.  
 PR 30-AUG-2000; 2000US-0227009.  
 PR 01-SEP-2000; 2000US-0229287.  
 PR 01-SEP-2000; 2000US-0229343.  
 PR 01-SEP-2000; 2000US-0229344.  
 PR 05-SEP-2000; 2000US-0229509.

PR 05-SEP-2000; 2000US-0229513.  
 PR 06-SEP-2000; 2000US-0230437.  
 PR 06-SEP-2000; 2000US-0230438.  
 PR 08-SEP-2000; 2000US-0231242.  
 PR 08-SEP-2000; 2000US-0231243.  
 PR 08-SEP-2000; 2000US-0231413.  
 PR 08-SEP-2000; 2000US-0231414.  
 PR 08-SEP-2000; 2000US-0232080.  
 PR 08-SEP-2000; 2000US-0232081.  
 PR 12-SEP-2000; 2000US-0231968.  
 PR 12-SEP-2000; 2000US-0232397.  
 PR 14-SEP-2000; 2000US-0232398.  
 PR 14-SEP-2000; 2000US-0232399.  
 PR 14-SEP-2000; 2000US-0232400.  
 PR 14-SEP-2000; 2000US-0232401.  
 PR 14-SEP-2000; 2000US-0233063.  
 PR 14-SEP-2000; 2000US-0233064.  
 PR 14-SEP-2000; 2000US-0233105.  
 PR 21-SEP-2000; 2000US-0234223.  
 PR 21-SEP-2000; 2000US-0234274.  
 PR 25-SEP-2000; 2000US-0234997.  
 PR 25-SEP-2000; 2000US-0234998.  
 PR 26-SEP-2000; 2000US-0235484.  
 PR 27-SEP-2000; 2000US-0235834.  
 PR 27-SEP-2000; 2000US-0235836.  
 PR 29-SEP-2000; 2000US-0236327.  
 PR 29-SEP-2000; 2000US-0236368.  
 PR 29-SEP-2000; 2000US-0236639.  
 PR 29-SEP-2000; 2000US-023670.  
 PR 02-OCT-2000; 2000US-0236802.  
 PR 02-OCT-2000; 2000US-0237037.  
 PR 02-OCT-2000; 2000US-0237038.  
 PR 02-OCT-2000; 2000US-0237039.  
 PR 02-OCT-2000; 2000US-0237040.  
 PR 13-OCT-2000; 2000US-0239935.  
 PR 13-OCT-2000; 2000US-0239937.  
 PR 20-OCT-2000; 2000US-0240960.  
 PR 20-OCT-2000; 2000US-0241121.  
 PR 20-OCT-2000; 2000US-0241785.  
 PR 20-OCT-2000; 2000US-0241786.  
 PR 20-OCT-2000; 2000US-0241787.  
 PR 20-OCT-2000; 2000US-0241808.  
 PR 20-OCT-2000; 2000US-0241809.  
 PR 01-NOV-2000; 2000US-0241826.  
 PR 01-NOV-2000; 2000US-0241826.  
 PR 08-NOV-2000; 2000US-0246474.  
 PR 08-NOV-2000; 2000US-0246475.  
 PR 08-NOV-2000; 2000US-0246476.  
 PR 08-NOV-2000; 2000US-0246477.  
 PR 08-NOV-2000; 2000US-0246478.  
 PR 08-NOV-2000; 2000US-0246523.  
 PR 08-NOV-2000; 2000US-0246524.  
 PR 08-NOV-2000; 2000US-0246525.  
 PR 08-NOV-2000; 2000US-0246527.  
 PR 08-NOV-2000; 2000US-0246528.  
 PR 08-NOV-2000; 2000US-0246532.  
 PR 08-NOV-2000; 2000US-0246609.  
 PR 08-NOV-2000; 2000US-0246610.  
 PR 08-NOV-2000; 2000US-0246611.  
 PR 08-NOV-2000; 2000US-0246613.  
 PR 08-NOV-2000; 2000US-0246617.  
 PR 17-NOV-2000; 2000US-0249207.  
 PR 17-NOV-2000; 2000US-0249208.  
 PR 17-NOV-2000; 2000US-0249209.  
 PR 17-NOV-2000; 2000US-0249210.  
 PR 17-NOV-2000; 2000US-0249211.  
 PR 17-NOV-2000; 2000US-0249212.  
 PR 17-NOV-2000; 2000US-0249213.  
 PR 17-NOV-2000; 2000US-0249214.  
 PR 17-NOV-2000; 2000US-0249215.  
 PR 17-NOV-2000; 2000US-0249216.

PR 17-NOV-2000; 2000US-0249217.  
 PR 17-NOV-2000; 2000US-0249218.  
 PR 17-NOV-2000; 2000US-0249244.  
 PR 17-NOV-2000; 2000US-0249245.  
 PR 17-NOV-2000; 2000US-0249264.  
 PR 17-NOV-2000; 2000US-0249265.  
 PR 17-NOV-2000; 2000US-0249287.  
 PR 17-NOV-2000; 2000US-0249299.  
 PR 01-DEC-2000; 2000US-0249300.  
 PR 01-DEC-2000; 2000US-0250310.  
 PR 05-DEC-2000; 2000US-0251030.  
 PR 05-DEC-2000; 2000US-0251988.  
 PR 05-DEC-2000; 2000US-0256179.  
 PR 06-DEC-2000; 2000US-0251479.  
 PR 08-DEC-2000; 2000US-0251856.  
 PR 08-DEC-2000; 2000US-0251868.  
 PR 08-DEC-2000; 2000US-0251869.  
 PR 08-DEC-2000; 2000US-0251989.  
 PR 08-DEC-2000; 2000US-0251990.  
 PR 11-DEC-2000; 2000US-0254097.  
 PR 05-JAN-2001; 2001US-0255678.  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 XX Rosen CA, Barash SC, Ruben SM;  
 XX WPI; 2001-488783/53.  
 DR N-PSDB; AAS26053.  
 XX  
 PT New nucleic acid molecules encoding 461 human secreted proteins for  
 PT diagnosing, preventing, treating or ameliorating medical conditions and  
 PT used as food additives or preservatives -  
 XX  
 PS Claim 11; SEQ ID NO 1019; 980PP; English.  
 XX  
 CC The invention relates to isolated nucleic acid molecules and their  
 CC encoded secreted proteins. The nucleic acids and proteins are used to  
 CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,  
 CC rabbits, goats, horses, cats, dogs, chickens or sheep. They  
 CC are also used in diagnosing a pathological condition or susceptibility  
 CC to a pathological condition. Antibodies to the proteins can also  
 CC be used in alleviating symptoms associated with the disorders and in  
 CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked  
 CC immunosorbant assays (ELISA). Disorders which are diagnosed or treated  
 CC include autoimmune diseases e.g. rheumatoid arthritis,  
 CC hyperproliferative disorders e.g. neoplasms of the breast or liver,  
 CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders  
 CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.  
 CC Alzheimer's disease, infections caused by bacteria, viruses and fungi.  
 CC and ocular disorders e.g. corneal infection, and many other  
 CC disorders listed in the specification. The polypeptides can also  
 CC be used to aid wound healing and epithelial cell proliferation, to  
 CC prevent skin aging due to sunburn, to maintain organs before  
 CC transplantation, for supporting cell culture of primary tissues, to  
 CC regenerate tissues and in chemotaxis. The polypeptides can also be used  
 CC as a food additive or preservative to increase or decrease storage  
 CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,  
 CC minerals, cofactors and other nutritional components. The present  
 CC sequence represents a novel secreted protein of the invention.

Query Match Score 61; DB 22; Length 429;  
 Best Local Similarity 66.3%; Pred. No. 0.58;  
 Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 1 GVKGDKGNPGWQAP 15  
 Db 23 GAKGEKGKGNAGFPGLP 37

RESULT 24  
 ID AA40114

XX AC AA40114;  
 XX DT 03-JUN-1998 (first entry)  
 XX DE Human alpha-1(IV) collagen protein.  
 XX KW COL4A6 gene; type IV collagen protein; alpha-6(IV); alpha-1(IV);  
 XX Alports syndrome; leiomyomatosis; diagnosis; gene therapy; antibody.  
 XX OS Homo sapiens.

XX Location/Qualifiers  
 XX 163..171  
 FT Key  
 FT Region  
 FT /label= I  
 FT /note= "This region interrupts a collagenous repeat"  
 FT 224..225  
 FT Region  
 FT /label= II  
 FT /note= "This region interrupts a collagenous repeat"  
 FT 244..257  
 FT Region  
 FT /label= III  
 FT /note= "This region interrupts a collagenous repeat"  
 FT 273..274  
 FT Region  
 FT /label= IV  
 FT /note= "This region interrupts a collagenous repeat"  
 FT 335..337  
 FT Region  
 FT /label= V  
 FT /note= "This region interrupts a collagenous repeat"  
 FT 377..378  
 FT Region  
 FT /label= VI  
 FT /note= "This region interrupts a collagenous repeat"  
 FT 403..404  
 FT Region  
 FT /label= VII  
 FT /note= "This region interrupts a collagenous repeat"  
 FT 429..236  
 FT Region  
 FT /label= VIII  
 FT /note= "This region interrupts a collagenous repeat"  
 FT 467..473  
 FT Region  
 FT /label= IX  
 FT /note= "This region interrupts a collagenous repeat"  
 PN US5731192-A.  
 XX PD 24-MAR-1998.  
 XX PF 23-JUN-1995; 95US-0494168.  
 XX PR 27-AUG-1993; 93US-0112465.  
 XX PA (UYYA ) UNIV YALE.  
 XX PI Reeder ST, Zhou J;  
 XX DR WPI; 1998-216495/19.  
 XX  
 PT Nucleic acid encoding human alpha-6(IV) collagen - useful for, e.g.  
 PT diagnosis or gene therapy of Alport's disease  
 XX  
 PS Example 1; FIG 3; 43pp; English.  
 XX  
 CC This sequence is the human alpha-1(IV) collagen protein which is used  
 CC to analyse a novel human alpha-6(IV) collagen protein encoded by the  
 CC COL4A6 gene. The alpha-6(IV) protein can be used for diagnosis or gene  
 CC therapy of diseases associated with collagen type IV pathology,  
 CC especially Alport's syndrome and associated diffuse leiomyomatosis. The  
 CC polypeptide may also be used for generating monoclonal or polyclonal  
 CC antibodies having specificity for the alpha-6(IV) polypeptide especially  
 CC an antibody that is not crossreactive with other collagen proteins  
 CC including alpha-1(IV), alpha-2(IV) and alpha-5(IV) collagens.  
 XX Sequence 532 AA;

SQ Query Match Score 61; DB 19; Length 532;

Best Local Similarity 66.7%; Pred. No. 0.73;  
Matches 10; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 1 GVKGDKGNGWPGAP 15  
 ||||:|||||:|||  
 Db 293 GDKCEKGSPGFPGEPEP 307

RESULT 25  
 AAM23916 standard; Protein; 1690 AA.  
 XX

AC AAM23916;  
 XX  
 DT 12-OCT-2001 (first entry)  
 XX

DE Human EST encoded protein SEQ ID NO: 1441.  
 XX

KW Human; sheep; pig; cow; fruit fly; yeast; hamster; macaque; horse;  
 KW tomato; monkey; dog; sea urchin; expressed sequence tag; EST;  
 KW diagnostics; forensic test; gene mapping; genetic disorder;  
 KW biodiversity; gene therapy; nutrition.  
 XX

OS Homo sapiens.  
 XX

PN WO200154477-A2.  
 XX

PD 02-AUG-2001.  
 XX

PF 25-JAN-2001; 2001WO-US02687.  
 XX

PR 17-JUL-2000; 2000US-0617746.  
 PR 03-AUG-2000; 2000US-0631451.  
 PR 15-SEP-2000; 2000US-0663870.

PA (OSTE-) OSTEOMETER AS.  
 XX

PI Bonde M, Qvist P;  
 XX

DR WPT; 1995-131456/17.  
 XX

PT Assaying collagen fragments in body fluid by immunoassay - using  
 PT antibodies raised against synthetic peptide(s) contg. potential  
 PT crosslinking sites, to diagnose and monitor disorders of collagen  
 XX

PS Disclosure (Appendix A); Page 55; 87pp; English.  
 XX

CC Determination of collagen fragments in body fluids can be achieved  
 CC by immunoassay using antibodies directed against synthetic peptides  
 CC derived from collagen which contain sites of potential crosslinking.  
 CC The method is used to diagnose and monitor treatment of disorders of  
 CC collagen metabolism (degradation of type I collagen may indicate  
 CC osteoporosis; metastatic progression, Paget's disease,  
 CC hyperthyroidism or other conditions involving excessive bone  
 CC resorption; degredation of type II collagen may indicate rheumatoid  
 CC arthritis or osteoarthritis; and of type III collagen, vacuilitis  
 CC syndrome). The method can also be used to assess the toxicity of a  
 CC compound and to test drugs for their effect on collagen metabolism.  
 XX

SQ Sequence 1078 AA;

The present invention provides the protein and coding sequences of novel  
 proteins from a variety of organisms, including human, dog, cat, horse,  
 cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea  
 urchin and tomato. These were derived from expressed sequence tags (ESTs)  
 from the organism of interest. They can be used in diagnostics,  
 forensics, gene mapping, identification of mutations, to assess  
 biodiversity and for nutritional purposes. The present sequence is a  
 protein of the invention.

XX

SQ Sequence 1690 AA;

The present invention provides the protein and coding sequences of novel  
 proteins from a variety of organisms, including human, dog, cat, horse,  
 cow, pig, hamster, monkey, macaque, yeast, bacteria, fruit fly, sea  
 urchin and tomato. These were derived from expressed sequence tags (ESTs)  
 from the organism of interest. They can be used in diagnostics,  
 forensics, gene mapping, identification of mutations, to assess  
 biodiversity and for nutritional purposes. The present sequence is a  
 protein of the invention.

XX

RESULT 27  
 AV96125  
 ID AV96125 standard; Peptide; 1078 AA.  
 XX

AC AV96125;  
 XX

DT 19-DEC-2000 (first entry)  
 XX

DE Collagen type III alpha-1.  
 XX

KW Collagen type III; vasculitis syndrome; assay; diagnosis.  
 XX

OS Homo sapiens.  
 XX

PN US6110689-A.  
 XX

RESULT 26  
 AAR71704  
 ID AAR71704 standard; protein; 1078 AA.  
 XX

PD 29-AUG-2000.  
 XX PT - contains primer to amplify portions of Type III procollagen DNA  
 PF 04-NOV-1997; PT and detects mutation in standard procollagen DNA  
 XX PR Disclosure: Fig 1A-F; 44pp; English.  
 XX (OSTE-) OSTROMETER AS.  
 XX CC Example 1 describes the determination of the presence of a mutation  
 PA in the proalpha1(III) gene. Primers used in PCR are given in  
 XX CC AAQ30834-48.  
 PI Bonde M, Ovist P;  
 XX PT  
 DR WPI; 2000-586349/55.  
 XX PT Assaying type I collagen fragments for diagnosing osteoporosis in  
 PT postmenopausal woman, involves contacting body fluid with synthetic  
 PT collagen peptide and antibody and quantifying by competitive binding  
 XX assay -  
 PS disclosure; Column 46-51; 41pp; English.  
 XX  
 CC The present sequence is that of human type III collagen alpha-1.  
 CC The invention is based on the discovery of the presence of  
 CC particular collagen fragments in body fluids of patients compared  
 CC with those of healthy subjects. These fragments are generated  
 CC upon collagen degradation and are partly characterised by the  
 CC presence of potential sites for crosslinking. A method for  
 CC assaying collagen fragments in a body fluid sample is based on the  
 CC competitive binding to immunological binding partners of collagen  
 CC fragments in the sample and of synthetic peptides derived from  
 CC collagen and containing crosslinkable sites (see AAY96118-21). When  
 CC considering the degradation of type III collagen, the assay can be  
 CC used as a means of identifying the presence of vasculitis syndrome.  
 XX  
 SQ Sequence 1078 AA:  
 Query Match 65.2%; Score 60; DB 21; Length 1196;  
 Best Local Similarity 71.4%; Pred. No. 2.4; Mismatches 1; Indels 0; Gaps 0;  
 Matches 10; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
 Qy 1 GVKGDKGNPGWGA 14  
 ||||:|||||  
 Db 711 GVKGGERGSPGGPPGA 724  
 ||||:|||||  
  
 RESULT 28  
 AAR28316  
 ID AAR28916 standard; Protein; 1196 AA.  
 XX  
 AC AAR28916;  
 XX DT 24-MAR-1993 (first entry)  
 XX DE Type III procollagen (prior art).  
 XX KW Mutation; Pro-alpha1(III); primer; PCR.  
 XX OS Homo sapiens.  
 XX PN WO200175177-A2.  
 XX PN 11-OCT-2001.  
 XX PR 03-APR-2001; 2001WO-US10947.  
 XX PR 03-APR-2000; 2000US-194336P.  
 XX PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
 XX PI Morin PJ, Sherman-Baust CA, Pizer ES, Hough CD;  
 XX DR WPI; 2001-626450/72.  
 XX DR N-PSDB; ARA83117.  
 XX PT Detecting and identifying ovarian tumor, identifying increased risk for  
 PT developing ovarian cancer, and determining effectiveness of ovarian  
 PT cancer treatment, by measuring expression level of ovarian tumor marker  
 PT gene -  
 XX PS Claim 23; Page 114-117; 140pp; English.  
 XX  
 PA (UYJE-) UNTV JEFFERSON THOMAS.  
 XX  
 PA Kuvvalaemi SH, Prockop DJ, Tromp GC;  
 XX WPI; 1992-398878/48.  
 DR N-PSDB; ARAQ30849.  
 XX  
 PT Kit for detecting genetic pre-disposition for vascular aneurysms

PT - contains primer to amplify portions of Type III procollagen DNA  
 PT and detects mutation in standard procollagen DNA  
 PF Disclosure: Fig 1A-F; 44pp; English.  
 XX (OSTE-) OSTROMETER AS.  
 XX CC Example 1 describes the determination of the presence of a mutation  
 CC in the proalpha1(III) gene. Primers used in PCR are given in  
 CC AAQ30834-48.  
 PI Bonde M, Ovist P;  
 XX PT  
 DR WPI; 2000-586349/55.  
 XX PT Assaying type I collagen fragments for diagnosing osteoporosis in  
 PT postmenopausal woman, involves contacting body fluid with synthetic  
 PT collagen peptide and antibody and quantifying by competitive binding  
 XX assay -  
 PS disclosure; Column 46-51; 41pp; English.  
 XX  
 CC The present sequence is that of human type III collagen alpha-1.  
 CC The invention is based on the discovery of the presence of  
 CC particular collagen fragments in body fluids of patients compared  
 CC with those of healthy subjects. These fragments are generated  
 CC upon collagen degradation and are partly characterised by the  
 CC presence of potential sites for crosslinking. A method for  
 CC assaying collagen fragments in a body fluid sample is based on the  
 CC competitive binding to immunological binding partners of collagen  
 CC fragments in the sample and of synthetic peptides derived from  
 CC collagen and containing crosslinkable sites (see AAY96118-21). When  
 CC considering the degradation of type III collagen, the assay can be  
 CC used as a means of identifying the presence of vasculitis syndrome.  
 XX  
 SQ Sequence 1196 AA:  
 Query Match 65.2%; Score 60; DB 13; Length 1196;  
 Best Local Similarity 71.4%; Pred. No. 2.4; Mismatches 3; Indels 1; Gaps 0;  
 Matches 10; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
 Qy 1 GVKGDKGNPGWGA 14  
 ||||:|||||  
 Db 858 GVKGGERGSPGGPPGA 871  
 ||||:|||||

RESULT 29  
 ABB5091  
 ID ABB50291 standard; Protein; 1466 AA.  
 XX AC ABB50291;  
 XX DT 08-FEB-2002 (first entry)  
 XX DE Collagen type III alpha-1 ovarian tumour marker protein, SEQ ID NO:72.  
 XX KW Ovarian tumour marker gene; human; overexpression; upregulation;  
 KW epithelial tumour; cancer; diagnosis; prognosis; disease monitoring;  
 KW identification; serous cystadenoma; borderline serous tumour;  
 KW serous cystadenocarcinoma; mucinous cystadenocarcinoma;  
 KW mucinous cystadenoma; borderline mucinous tumour; endometrioid carcinoma;  
 KW adenofibroma; Brenner tumour; serial analysis of gene expression; SAGE;  
 KW immune response pathway; cell proliferation regulation; protein folding;  
 KW membrane localised; secreted; therapeutic target; cytostatic;  
 KW gene therapy; vaccine.  
 XX OS Homo sapiens.  
 XX PN WO200175177-A2.  
 XX PN 11-OCT-2001.  
 XX PR 03-APR-2001; 2001WO-US10947.  
 XX PR 03-APR-2000; 2000US-194336P.  
 XX PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
 XX PI Morin PJ, Sherman-Baust CA, Pizer ES, Hough CD;  
 XX DR WPI; 2001-626450/72.  
 XX DR N-PSDB; ARA83117.  
 XX PT Detecting and identifying ovarian tumor, identifying increased risk for  
 PT developing ovarian cancer, and determining effectiveness of ovarian  
 PT cancer treatment, by measuring expression level of ovarian tumor marker  
 PT gene -  
 XX PS Claim 23; Page 114-117; 140pp; English.  
 XX  
 PA (UYJE-) UNTV JEFFERSON THOMAS.  
 XX  
 PA Kuvvalaemi SH, Prockop DJ, Tromp GC;  
 XX WPI; 1992-398878/48.  
 DR N-PSDB; ARAQ30849.  
 XX  
 PT Kit for detecting genetic pre-disposition for vascular aneurysms

CC monitoring a patient in remission from ovarian cancer and in tests for  
 CC monitoring disease status in a patient being treated for ovarian cancer.  
 CC The methods can additionally be used to identify a particular tumour as  
 CC being an ovarian tumour (i.e., an epithelial ovarian tumour selected from  
 CC serous cystadenoma, borderline serous tumour, serous cystadenocarcinoma,  
 CC mucinous cystadenoma, borderline mucinous tumour, mucinous  
 CC cystadenocarcinoma, endometrioid carcinoma, undifferentiated carcinoma,  
 CC clear cell adenocarcinoma, cystadenofibroma, adenofibroma and Brenner  
 CC tumour. The ovarian tumour marker genes of the invention were identified  
 CC using SAGE (serial analysis of gene expression) and were found to be  
 CC overexpressed in a broad variety of ovarian epithelial tumour cells  
 relative to normal ovarian epithelial cells. The marker genes are  
 CC implicated in immune response pathways, in the regulation of cell  
 CC proliferation and in protein folding, and many of these are membrane-  
 CC localised or secreted. In addition to their use as diagnostic and  
 CC prognostic markers, the ovarian tumour marker genes or their encoded  
 CC proteins may be used as therapeutic targets for the treatment and  
 CC prevention of ovarian cancer. Sequences ABP50257-ABP50299 represent  
 CC proteins encoded by ovarian tumour marker genes of the invention.

SQ Sequence 1466 AA:  
 Best Local Similarity 65.2%; Score 60; DB 22; Length 1466;  
 Matches 11; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GVKGDKGNPGWPGA 14  
 Db 738 GPKGDKGPGGPGA 751

RESULT 30

AAE02533 AAE02533 standard; Protein; 1466 AA.

XX AAE02533;  
 XX DT 10-AUG-2001 (first entry)  
 XX DE Bovine alpha(I,II) collagen #1.  
 XX KW Bovine; alpha(I,II) collagen; gelatin; cytostatic; viral infection;  
 KW pharmaceutical; food industry; cosmetic; autoimmune disorder; vaccine;  
 KW medical; arterial sealant; bone graft; dermal implant; haemostat; cancer;  
 KW rheumatoid arthritis; beverage; photographic application.  
 OS Bos sp.  
 XX PN WO200134647-A2.  
 XX PD 17-MAY-2001.  
 XX PP 10-NOV-2000; 2000000-US30792.  
 XX PR 12-NOV-1999; 95US-0439058.  
 PR 10-NOV-2000; 200000-0439058.  
 XX PA (FIBR-) FIBROGEN INC.  
 XX PI Bell MP, Neff TB, Polarek JW, Seeley TW;  
 XX DR WPI; 2001-335911/35.  
 XX N-FSDB; AAD06574.

PT Novel isolated and purified bovine or porcine collagens and gelatins  
 PT useful in medical, pharmaceutical, food and cosmetic industries, as  
 PT vaccine, and for treating autoimmune disorders, infections and cancer  
 PT  
 XX Example 2; FIG 4; 168pp; English.  
 XX  
 CC The present sequence is bovine alpha(I,II) collagen. The present  
 CC invention relates to recombinant synthesis of collagens and gelatins

CC derived from animals. Collagen is useful in medical pharmaceutical,  
 CC food and cosmetic industries. Collagen is an important component of  
 CC arterial sealants, bone grafts, drug delivery system, dermal implants,  
 CC haemostats, and incontinence implants, and for treating autoimmune  
 CC disorders such as rheumatoid arthritis. Collagen is useful in food  
 CC products such as sausages casings, and in cosmetics or facial and skin  
 CC formulations for treating viral infections, autoimmune diseases and  
 CC cancer. Gelatin is useful in the manufacture or as a component of  
 CC various pharmaceutical and medical devices and products, in food and  
 CC beverage industries, in hair care and skin care products, as a glue or  
 CC adhesive in various manufacturing processes, as a light-sensitive coating  
 CC in various electronic devices, as photoresist base in photolithographic  
 CC processes, in printing and photographic applications, in laboratory  
 CC application, and as a component in various gels used for biochemical and  
 CC electrophoretic analysis, including enzymographic gels.

SQ Sequence 1466 AA:  
 Best Local Similarity 65.2%; Score 60; DB 22; Length 1466;  
 Matches 10; Conservative 71.4%; Pred. No. 3; Mismatches 3; Indels 1; Gaps 0; Gaps 0;

QY 1 GVKGDKGNPGWPGA 14  
 Db 858 GVKGERSPGGFGA 871

Search completed: November 1, 2002, 12:52:33  
 Job time : 29.5 secs

Page 1

Copyright (c) 1993 - 2002 CompuGen Ltd.

GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: November 1, 2002, 12:52:42 ; Search time 10.5 Seconds

Minimum DB seq length: 0 (without alignments)  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%

Total number of hits satisfying chosen parameters: 231628

Scoring table: BiCSUM62  
Gapext 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Database : Issued\_Patents\_AA:\*

- 1: /cgn2\_6/pctodata/1/1aa/5A\_\_COMB.pep:\*
- 2: /cgn2\_6/pctodata/1/1aa/5B\_\_COMB.pep:\*
- 3: /cgn2\_6/pctodata/1/1aa/6A\_\_COMB.pep:\*
- 4: /cgn2\_6/pctodata/1/1aa/6B\_\_COMB.pep:\*
- 5: /cgn2\_6/pctodata/1/1aa/PCTUS.COMB.pep:\*
- 6: /cgn2\_6/pctodata/1/1aa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

### SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	92	100.0	15	US-08-394-748A-1
2	92	100.0	15	US-08-394-748A-1
3	92	100.0	15	US-08-394-748A-1
4	92	100.0	24	US-08-534-342-3
5	92	100.0	24	US-08-675-140-5
6	92	100.0	39	US-08-534-342-2
7	92	100.0	39	US-08-675-140-2
8	66	71.7	39	US-08-534-342-1
9	66	71.7	39	US-08-534-342-6
10	66	71.7	39	US-08-675-140-5
11	66	71.7	39	US-08-675-140-6
12	61	66.3	532	US-08-494-168-9
13	60	65.2	1057	US-08-931-820-4
14	60	65.2	1078	US-08-963-825-21
15	60	65.2	1078	US-08-570-573-21
16	60	65.2	1078	US-08-548-608-21
17	59	64.1	471	US-08-399-89-24
18	59	64.1	471	US-09-167-364-24
19	59	64.1	471	US-09-439-897-2
20	56	60.9	489	US-08-794-795-7
21	56	60.9	489	US-08-249-200-7
22	56	60.9	518	US-08-392-367B-2
23	56	60.9	518	US-08-489-89-24
24	56	60.9	546	US-08-494-168-10
25	55	59.8	171	US-09-011-735-2
26	55	59.8	351	US-09-011-735-1
27	55	59.8	1694	US-08-494-168-2

### ALIGNMENTS

RESULT 1  
Sequence 1, Application US/08394748A

GENERAL INFORMATION:

APPLICANT: Skubitz, Amy P.N.  
APPLICANT: Furcht, Leo T.  
APPLICANT: Balles, Mark  
APPLICANT: Grederson, Dale S.  
APPLICANT: Agarwal, Anita  
APPLICANT: Wright, Martha M.  
APPLICANT: Murali, Shobhana

TITLE OF INVENTION: Method for Treating Condition Using Polypeptides

NUMBER OF SEQUENCES: 16

CORRESPONDENCE ADDRESS:  
ADDRESSEE: Merchant & Gould  
STREET: 3100 No. 6013628west Center  
CITY: Minneapolis  
STATE: MN  
COUNTRY: USA  
ZIP: 55402

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPILER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.0

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/394,748A

APPLICATION NUMBER: US/08/394,748A

FILING DATE: 27-FEB-1995

CLASSIFICATION: 514

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/203,458

FILING DATE: 28-FEB-1994

ATTORNEY/AGENT INFORMATION:  
NAME: Carter, Charles G.

REGISTRATION NUMBER: 35,093

REFERENCE/DOCKET NUMBER: 600,307US01

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 612-332-5300  
TELEFAX: 612-332-9081

INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 15 amino acids

TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: Peptide  
ORIGINAL SOURCE:  
STRAIN: IV-H1

US-08-394-748A-1

```

Sequence 38, Appl
Sequence 2, Appl
Sequence 2, Appl
Sequence 6, Appl
Sequence 6, Appl
Sequence 8, Appl
Sequence 10, Appl
Sequence 10, Appl
Sequence 2, Appl
Sequence 1, Appl
Sequence 3, Appl
Sequence 6, Appl
Sequence 20, Appl
Sequence 1, Appl
Sequence 20, Appl
Sequence 20, Appl
Sequence 12, Appl
Sequence 12, Appl

```

Query Match 100.0%; Score 92; DB 3; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 2.1e-07; Indels 0; Gaps 0;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GVKGDKGNPGWPGAP 15  
 Db 1 GVKGDKGNPGWPGAP 15

**RESULT 2**  
 US-08-702-254A-1  
 Sequence 1, Application US/08702254A  
 Patent No. 609683

GENERAL INFORMATION:  
 APPLICANT: FIELDS, GREGG B.  
 TITLE OF INVENTION: SELF-ASSEMBLING AMPHIPHILES FOR CONSTRUCTION OF PEPTIDE SECONDARY STRUCTURES  
 NUMBER OF SEQUENCES: 3  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: MUETTING, RAASCH & GEBHARDT, P.A.  
 STREET: 119 NORTH FOURTH STREET  
 CITY: MINNEAPOLIS  
 STATE: MINNESOTA  
 COUNTRY: USA  
 ZIP: 55401

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/702,254A  
 FILING DATE: 23-AUG-1996  
 CLASSIFICATION: 530

ATTORNEY/AGENT INFORMATION:  
 NAME: MUETTING, ANN M.  
 REGISTRATION NUMBER: 33,977  
 REFERENCE/DOCKET NUMBER: 1110.00310101  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 612-305-1217  
 TELEX/FAX: 612-305-1228

INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 15 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: peptide  
 ;US-08-702-254A-1

Query Match 100.0%; Score 92; DB 5; Length 15;  
 Best Local Similarity 100.0%; Pred. No. 2.1e-07; Indels 0; Gaps 0;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GVKGDKGNPGWPGAP 15  
 Db 1 GVKGDKGNPGWPGAP 15

**RESULT 4**  
 US-08-534-342-3  
 Sequence 3, Application US/08534342  
 Patent No. 5576419

GENERAL INFORMATION:  
 APPLICANT: Fields, Gregg B.  
 TITLE OF INVENTION: Mild Solid-Phase Synthesis of Aligned, Numbered Sequences  
 NUMBER OF SEQUENCES: 13  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Merchant & Gould  
 STREET: 3100 No. 5576419west Center  
 CITY: Minneapolis  
 STATE: MN  
 COUNTRY: USA  
 ZIP: 55402

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/534,342  
 FILING DATE: 30-JUN-1993  
 CLASSIFICATION: 530

PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/085,633  
 FILING DATE: 30-JUN-1993

ATTORNEY/AGENT INFORMATION:  
 NAME: Mueting, Ann M.  
 REGISTRATION NUMBER: 33,977  
 REFERENCE/DOCKET NUMBER: 600.253-US-01  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 612-332-5300  
 TELEX/FAX: 612-332-9081

INFORMATION FOR SEQ ID NO: 3:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 24 amino acids  
 TYPE: amino acid  
 TOPOLOGY: unknown

US-08-534-342-3  
 Query Match 100.0%; Score 92; DB 1; Length 24;  
 Best Local Similarity 100.0%; Pred. No. 3.3e-07; Indels 0; Gaps 0;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GVKGDKGNPGWPGAP 15  
 ;US-08-534-342-3

Db 10 GVKGDKGNGPGWPAGP 24

CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/534, 342  
 FILING DATE:  
 CLASSIFICATION: 530  
 PRIORITY APPLICATION DATA:  
 APPLICATION NUMBER: US 08/085, 633  
 FILING DATE: 30-JUN-1993  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Muetling, Ann M.  
 REGISTRATION NUMBER: 33, 977  
 REFERENCE/DOCKET NUMBER: 600 253-US-01  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 612-332-5300  
 TELEFAX: 612-332-9081  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 39 amino acids  
 TYPE: amino acid  
 LENGTH: 39 amino acids  
 TOPOGY: unknown

US-08-534-342-2

Query Match 100.0%; Score 92; DB 1; Length 39;  
 Best Local Similarity 100.0%; Pred. No. 5.6e-07;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 7  
 US-08-675-140-2

Sequence 2, Application US/08675140  
 Patent No. 572643

GENERAL INFORMATION:  
 APPLICANT: Fields, Gregg B.  
 TITLE OF INVENTION: Mild Solid-Phase Synthesis of Aligned,  
 NUMBER OF SEQUENCES: 13  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Merchant & Gould  
 STREET: 3100 No. 5576419west Center  
 CITY: Minneapolis  
 STATE: MN  
 COUNTRY: USA  
 ZIP: 55402

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25

ATTORNEY/AGENT INFORMATION:  
 NAME: Muetling, Ann M.  
 REGISTRATION NUMBER: 33, 977  
 REFERENCE/DOCKET NUMBER: 600 253-US-01  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 612-332-5300  
 TELEFAX: 612-332-9081  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 39 amino acids  
 TYPE: amino acid  
 LENGTH: 39 amino acids  
 TOPOGY: unknown

US-08-675-140-3

Query Match 100.0%; Score 92; DB 1; Length 24;  
 Best Local Similarity 100.0%; Pred. No. 3.3e-07;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 6  
 US-08-534-342-2

Sequence 2, Application US/08534342  
 Patent No. 5576419

GENERAL INFORMATION:  
 APPLICANT: Fields, Gregg B.  
 TITLE OF INVENTION: Mild Solid-Phase Synthesis of Aligned,  
 NUMBER OF SEQUENCES: 13  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Merchant & Gould  
 STREET: 3100 No. 5576419west Center  
 CITY: Minneapolis  
 STATE: MN  
 COUNTRY: USA  
 ZIP: 55402

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/675, 140  
 FILING DATE: 03-JUL-1996  
 CLASSIFICATION: 525  
 PRIORITY APPLICATION DATA:  
 APPLICATION NUMBER: US 08/085, 633  
 FILING DATE: 30-JUN-1993  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Muetling, Ann M.  
 REGISTRATION NUMBER: 33, 977  
 REFERENCE/DOCKET NUMBER: 600 253-US-01  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 612-332-5300  
 TELEFAX: 612-332-9081  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 39 amino acids  
 TYPE: amino acid  
 LENGTH: 39 amino acids  
 TOPOGY: unknown

US-08-675-140-2

Query Match 100.0%; Score 92; DB 1; Length 39;  
 Best Local Similarity 100.0%; Pred. No. 5.6e-07;

**RESULT 8**  
 US-08-534-342-5  
 ; Sequence 5, Application US/08534342  
 ; Patent No. 5576419  
 GENERAL INFORMATION:  
 APPLICANT: Fields, Gregg B.  
 TITLE OF INVENTION: Mild Solid-Phase Synthesis of Aligned,  
 NUMBER OF SEQUENCES: 13  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Merchant & Gould  
 STREET: 3100 NO. 5576419west Center  
 CITY: Minneapolis  
 STATE: MN  
 COUNTRY: USA  
 ZIP: 55402  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/534,342  
 FILING DATE: 30-JUN-1993  
 CLASSIFICATION: 530  
 PRIORITY APPLICATION DATA:  
 COMPUTER NUMBER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/534,342  
 FILING DATE:  
 CLASSIFICATION: 530  
 PRIORITY APPLICATION DATA:  
 COMPUTER NUMBER: US 08/085,633  
 FILING DATE: 30-JUN-1993  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Muetting, Ann M.  
 REGISTRATION NUMBER: 33,977  
 REFERENCE/DOCKET NUMBER: 600.253-US-01  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 612-332-5300  
 TELEFAX: 612-332-9081  
 INFORMATION FOR SEQ ID NO: 5:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 39 amino acids  
 TYPE: amino acid  
 TOPOLOGY: unknown  
 US-08-534-342-6  
 Query Match 71.7%; Score 66; DB 1; Length 39;  
 Best Local Similarity 73.3%; Pred. No. 0 003; Mismatches 0;  
 Matches 11; Conservative 0; Gaps 0;  
 Indels 0; Length 39;  
 Qy 1 GVKGDKGNPGWPGAP 15  
 Db 25 GVXGXGXPWPGAP 39  
 RESULT 9  
 US-08-534-342-6  
 ; Sequence 6, Application US/08534342  
 ; Patent No. 5576419  
 GENERAL INFORMATION:  
 APPLICANT: Fields, Gregg B.  
 TITLE OF INVENTION: Mild Solid-Phase Synthesis of Aligned,  
 NUMBER OF SEQUENCES: 13  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Merchant & Gould  
 STREET: 3100 NO. 5576419west Center  
 CITY: Minneapolis  
 STATE: MN  
 COUNTRY: USA  
 ZIP: 55402  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/675,140  
 FILING DATE: 03-JUL-1996  
 CLASSIFICATION: 525  
 PRIORITY APPLICATION DATA:  
 COMPUTER NUMBER: US 08/085,633  
 FILING DATE: 30-JUN-1993  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Muetting, Ann M.  
 REGISTRATION NUMBER: 33,977  
 REFERENCE/DOCKET NUMBER: 600.253-US-01  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 612-332-5300  
 INFORMATION FOR SEQ ID NO: 5:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 39 amino acids  
 TYPE: amino acid  
 TOPOLOGY: unknown  
 US-08-534-342-5  
 Query Match 71.7%; Score 66; DB 1; Length 39;  
 Best Local Similarity 73.3%; Pred. No. 0 003; Mismatches 0;  
 Matches 11; Conservative 0; Gaps 0;  
 Indels 0; Length 39;  
 Qy 1 GVKGDKGNPGWPGAP 15  
 Db 25 GVXGXGXPWPGAP 39  
 RESULT 10  
 US-08-675-140-5  
 ; Sequence 5, Application US/08675140  
 ; Patent No. 5726243  
 GENERAL INFORMATION:  
 APPLICANT: Fields, Gregg B.  
 TITLE OF INVENTION: Mild Solid-Phase Synthesis of Aligned,  
 NUMBER OF SEQUENCES: 13  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Merchant & Gould  
 STREET: 3100 NO. 5726243west Center  
 CITY: Minneapolis  
 STATE: MN  
 COUNTRY: USA  
 ZIP: 55402  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/675,140  
 FILING DATE: 03-JUL-1996  
 CLASSIFICATION: 525  
 PRIORITY APPLICATION DATA:  
 COMPUTER NUMBER: US 08/085,633  
 FILING DATE: 30-JUN-1993  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Muetting, Ann M.  
 REGISTRATION NUMBER: 33,977  
 REFERENCE/DOCKET NUMBER: 600.253-US-01  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 612-332-5300  
 INFORMATION FOR SEQ ID NO: 5:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 39 amino acids  
 TYPE: amino acid  
 TOPOLOGY: unknown  
 US-08-534-342-6  
 Query Match 71.7%; Score 66; DB 1; Length 39;  
 Best Local Similarity 73.3%; Pred. No. 0 003; Mismatches 0;  
 Matches 11; Conservative 0; Gaps 0;  
 Indels 0; Length 39;

US-08-675-140-6

Query Match 71.7%; Score 66; DB 1; Length 39;

Best Local Similarity 73.3%; Pred. No. 0.003; 0; Mismatches

Matches 11; Conservative 0; Indels 4; Gaps 0;

Qy 1 GVKGDKGNPGPGAP 15

Db 25 GVGXGKXGPGPAG 39

RESULT 11

US-08-675-140-6

Sequence 6, Application US/08675140

Patent No. 5726243

GENERAL INFORMATION:

APPLICANT: Fields, Gregg B. TITLE OF INVENTION: Mild Solid-Phase Synthesis of Aligned,

NUMBER OF SEQUENCES: 13

CORRESPONDENCE ADDRESS:

ADDRESSEE: Merchant &amp; Gould

STREET: 3100 No. 5726243west Center

CITY: Minneapolis

STATE: MN

COUNTRY: USA

ZIP: 55402

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/675,140

FILING DATE: 03-JUL-1996

CLASSIFICATION: 525

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/085,633

FILING DATE: 30-JUN-1993

ATTORNEY / AGENT INFORMATION:

NAME: Muetting, Ann M.

REGISTRATION NUMBER: 33,977

TELECOMMUNICATION INFORMATION:

TELEPHONE: 612-332-5300

TELEFAX: 612-332-9081

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 39 amino acids

TYPE: amino acid

TOPOLOGY: unknown

US-08-675-140-6

Query Match 71.7%; Score 66; DB 1; Length 39;

Best Local Similarity 73.3%; Pred. No. 0.003; 0; Mismatches

Matches 11; Conservative 0; Indels 4; Gaps 0;

Qy 1 GVKGDKGNPGPGAP 15

Db 25 GVGXGKXGPGPAG 39

RESULT 12

US-08-494-168-9

Sequence 9, Application US/08494168

Patient No. 5731192

GENERAL INFORMATION:

APPLICANT: Reeder, Stephen T.

TITLE OF INVENTION: Collagen COL4A6: Gene, Protein and Method

NUMBER OF SEQUENCES: 10

CORRESPONDENCE ADDRESS:

ADDRESSEE: Foley &amp; Lardner

STREET: 3000 K Street, N.W., Suite 500

CITY: Washington, D.C.

COUNTRY: USA

ZIP: 20007-5109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/494,168

FILING DATE: 4/35

PRIORITY APPLICATION DATA:

APPLICATION NUMBER: US/08/112,465

FILING DATE: 27-AUG-1993

ATTORNEY/AGENT INFORMATION:

NAME: SAME, Bernhard D.

REGISTRATION NUMBER: 28,665

REFERENCE/DOCKET NUMBER: 40397/04/BABR

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202)672-5300

TELEFAX: (202)672-5399

TELEX: 904135

INFORMATION FOR SEQ ID NO: 9:

SEQUENCE CHARACTERISTICS:

LENGTH: 53 amino acids

TYPE: amino acid

TOPOLOGY: linear

US-08-494-168-9

Query Match 66.3%; Score 61; DB 1; Length 532;

Best Local Similarity 66.7%; Pred. No. 0.24; 3; Mismatches

Matches 10; Conservative 3; Indels 2; Gaps 0;

Qy 1 GVKGDKGNPGPGAP 15

Db 293 GDGKEKGSPGFPGEP 307

RESULT 13

US-08-931-820-4

Sequence 4, Application US/08931820

PATENT NO. 6010863

GENERAL INFORMATION:

APPLICANT: Assay for collagen degradation

TITLE OF INVENTION: Assay for collagen degradation

NUMBER OF SEQUENCES: 4

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/931,820

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: EP 96202596.1

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 1057 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: NO

ORIGINAL SOURCE:

ORGANISM: Homo sapiens

TISSUE TYPE: Collagen type III

FEATURE:

NAME/KEY: Modified-site

LOCATION: 1055  
; OTHER INFORMATION: /label= Modified  
; OTHER INFORMATION: /note= "Ala may be Pro"  
; US-08-931-820-4

Query Match: Best Local Similarity 71.4%; Score 60; DB 3; Length 1057;  
; Matches 10; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 GVKDKGNGWPGKA 14  
; ||||:|||||  
Db 710 GVKEERGSPPGPA 723

RESULT 14  
US-08-903-825-21  
; Sequence 21, Application US/08963825  
; Patent No. 6110639

GENERAL INFORMATION:  
; APPLICANT: Qvist, Per  
; COMPUTER: Bond, Martin  
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments  
; IN BODY FLUIDS, A Test Kit and Means for Carrying Out the  
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of  
; TITLE OF INVENTION: Disorders Associated with the Metabolism of  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Darby & Darby PC  
; STREET: 805 Third Avenue  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10022

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/963,825  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/187,319  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Gogoris, Adda C  
; REGISTRATION NUMBER: 29,714  
; REFERENCE/DOCKET NUMBER: 4305/08701  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-527-7700  
; TELEFAX: 212-753-6237  
; TELEX: 236687

INFORMATION FOR SEQ ID NO: 21:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1078 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; ORIGINAL SOURCE:  
; ORGANISM: Homo sapiens  
; IMMEDIATE SOURCE:  
; CLONE: COLLAGEN ALPHA 1 (III)  
; US-09-570-573-21

Query Match: Best Local Similarity 71.4%; Score 60; DB 4; Length 1078;  
; Matches 10; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 GVKDKGNGWPGKA 14  
; ||||:|||||  
Db 711 GVKEERGSPPGPA 724

RESULT 15  
US-09-570-573-21  
; Sequence 21, Application US/09570573  
; Patent No. 6342361

GENERAL INFORMATION:  
; APPLICANT: Qvist, Per  
; COMPUTER: Bond, Martin  
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments  
; IN BODY FLUIDS, A Test Kit and Means for Carrying Out the  
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of  
; TITLE OF INVENTION: Disorders Associated with the Metabolism of  
; NUMBER OF SEQUENCES: 21

Query Match: Best Local Similarity 71.4%; Score 60; DB 3; Length 1078;  
; Matches 10; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 GVKDKGNGWPGKA 14  
; ||||:|||||  
Db 711 GVKEERGSPPGPA 724

RESULT 16  
US-09-548-608-21  
; Sequence 21, Application US/09548608  
; Patent No. 635542

GENERAL INFORMATION:  
; APPLICANT: Qvist, Per  
; COMPUTER: Bond, Martin  
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments  
; IN BODY FLUIDS, A Test Kit and Means for Carrying Out the  
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of  
; TITLE OF INVENTION: Disorders Associated with the Metabolism of  
; NUMBER OF SEQUENCES: 21

CORRESPONDENCE ADDRESS:

ADDRESSEE: Darby & Darby PC

STREET: 805 Third Avenue

CITY: New York

STATE: New York

COUNTRY: USA

ZIP: 10022

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/548,608

FILING DATE:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/187,319

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Goporis, Adda C

REGISTRATION NUMBER: 29,714

REFERENCE/DOCKET NUMBER: 4305/08701

SEQUENCE CHARACTERISTICS:

TELEPHONE: 212-527-7700

TELEFAX: 212-553-6237

TELEX: 236687

IMFORMATION FOR SEQ ID NO: 21:

SEQUENCE CHARACTERISTICS:

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

ORIGINAL SOURCE:

ORGANISM: Homo sapiens

IMMEDIATE SOURCE:

CLONE: COLLAGEN ALPHA 1 (III)

US-09-548-608-21

Query Match 65.2%; Score 60; DB 4; Length 1078;

Best Local Similarity 71.4%; Pred. No. 0.69;

Matches 10; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GVGDKGNPGWPGCA 14

Db 711 GVGGERGSPGGFGA 724

RESULT 17

US-08-399-889-24

Sequence 24, Application US/08399889B

Patient No. 597120

GENERAL INFORMATION:

APPLICANT: Readers, Stephen T

APPLICANT: Morrison, Karen E

APPLICANT: Hudson, Billy G

TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides

FILE REFERENCE: 951263A

CURRENT APPLICATION NUMBER: US/08/399,889

CURRENT FILING DATE: 1995-03-07

EARLIER APPLICATION NUMBER: 07/621091

EARLIER FILING DATE: 1990-11-30

NUMBER OF SEQ ID NOS: 25

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 24

LENGTH: 471

TYPE: PRT

ORGANISM: Calf

US-08-399-889-24

Qy 1 GVKGDKGNGPWPAGP 15

Db 116 GMKGKKGNSGFPGPP 130

RESULT 18

US-09-167-364-24

Sequence 24, Application US/09167364

Patient No. 6007980

GENERAL INFORMATION:

APPLICANT: Reeders, Stephen T

APPLICANT: Morrison, Karen E

APPLICANT: Hudson, Billy G

TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides

FILE REFERENCE: 951263B

CURRENT APPLICATION NUMBER: US/09/167,364

CURRENT FILING DATE: 1998-10-07

EARLIER APPLICATION NUMBER: 08/399889

EARLIER FILING DATE: 1995-03-07

NUMBER OF SEQ ID NOS: 25

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 24

LENGTH: 471

TYPE: PRT

ORGANISM: Calf

US-09-167-364-24

RESULT 19

US-09-39-897-2

Sequence 2, Application US/09439897

Patient No. 6277558

GENERAL INFORMATION:

APPLICANT: Hudson, Billy G

TITLE OF INVENTION: Alpha-3 Chain Type IV Collagen Polypeptides

FILE REFERENCE: 95-1263-C

CURRENT APPLICATION NUMBER: US/09/439,897

CURRENT FILING DATE: 1999-11-12

NUMBER OF SEQ ID NOS: 65

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 2

LENGTH: 471

TYPE: PRT

ORGANISM: Bos taurus

US-09-39-897-2

Query Match 64.1%; Score 59; DB 4; Length 471;

Best Local Similarity 66.7%; Pred. No. 0.4;

Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GVKGDKGNGPWPAGP 15

Db 116 GMKGKKGNSGFPGPP 130

RESULT 20

US-08-794-795-7

Sequence 7, Application US/08794795

Patient No. 591676

GENERAL INFORMATION:

APPLICANT: Elshourbagy, Nabil

APPLICANT: Adamou, John

APPLICANT: Gross, Mitchell

APPLICANT: Lysko, Paul

TITLE OF INVENTION: Human Macro Scavenger Rec  
 TITLE OF INVENTION: eptor  
 NUMBER OF SEQUENCES: 9  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Smithkline Beecham Corporation  
 STREET: 709 Sweetbriar Road  
 CITY: King of Prussia  
 STATE: PA  
 COUNTRY: USA  
 ZIP: 19406  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette  
 COMPUTER: IBM Compatible  
 OPERATING SYSTEM: DOS  
 SOFTWARE: FastSeq for Windows Version 2.0  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/794,795  
 FILING DATE: 04-FEB-1997  
 CLASSIFICATION: 435  
 PRIORITY APPLICATION DATA:  
 APPLICATION NUMBER: ATG50009P  
 FILING DATE: 22-MAY-1996  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Han, William T  
 REGISTRATION NUMBER: 34,344  
 REFERENCE/DOCKET NUMBER: ATG50009  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 610-270-5219  
 TELEX: 610-270-4026  
 INFORMATION FOR SEQ ID NO: 7:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 489 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 ; US-08-794-795-7

Query Match 60.9%; Score 56; DB 2; Length 489;  
 Best Local Similarity 66.7%; Pred. No. 1.1; Mismatches 10; Conservative 1; Indels 4; Gaps 0; Gaps 0;  
 Matches 10; Mismatches 1; Indels 4; Gaps 0;

Qy 1 GVKGDKGNGWPGAP 15  
 |||||:|||\_|  
 Db 265 GVKGDQGKPGVQGV 279

RESULT 22-367B-2  
 US-08-392-367B-2  
 ; Sequence 2, Application US/08392367B  
 ; Patent No. 5691197  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Tryggvason, Karl  
 ; APPLICANT: Elomaa, Outi  
 ; APPLICANT: Kangas, Maari  
 ; TITLE OF INVENTION: An Insolated DNA Sequence For a  
 ; Patent No. 5691197  
 ; TITLE OF INVENTION: NO. 5691197e1 Macrophage Receptor with  
 ; TITLE OF INVENTION: a Collagenous Domain and the  
 ; TITLE OF INVENTION: Polypeptide Chain Encoded by  
 ; TITLE OF INVENTION: such a Sequence  
 ; NUMBER OF SEQUENCES: 2  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Fay, Sharpe, Beall, Fagan,  
 ; ADDRESS: Minnich & McKee  
 ; STREET: 1100 Superior Avenue  
 ; STREET: Suite 700  
 ; CITY: Cleveland  
 ; STATE: Ohio  
 ; COUNTRY: U.S.A.  
 ; ZIP: 44114-2518  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette, 3.50 inch,  
 MEDIUM TYPE: 720 Kb storable  
 COMPUTER: IBM PS/2, Model 35 SX  
 OPERATING SYSTEM: DOS 5.0  
 SOFTWARE: Word Perfect 5.1  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/392,367B  
 FILING DATE:  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Minnich, Richard J.  
 REGISTRATION NUMBER: 24,175  
 REFERENCE/DOCKET NUMBER: TRV 2 009  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (216) 861-5582

APPLICATION NUMBER: US/09/249,200  
 ;

TELEFAX: (216) 241-1666  
 TELEX: (216) 980162  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 518 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: Single  
 TOPOLOGY: Linear

; US-08-392-367B-2  
 Query Match, Best Local Similarity 60.9%; Score 56; DB 1; Length 518;  
 Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

RESULT 23  
 US-08-893-467A-2  
 Sequence 2, Application US/08893467A

PATENT NO. 6063901  
 GENERAL INFORMATION:  
 APPLICANT: Tryggason, Karl  
 APPLICANT: Elomaa, Outi  
 APPLICANT: Kangas, Marit  
 TITLE OF INVENTION: An Insolated DNA Sequence For a  
 Patent No. 6063901  
 TITLE OF INVENTION: NO. 6063901el Macrophage Receptor with  
 a Collagenous Domain and the  
 Polypeptide Chain Encoded by  
 NUMBER OF SEQUENCES: 2

CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Fay Sharpe, Beall, Fagan,  
 STREET: 1100 Superior Avenue  
 CITY: Cleveland  
 STATE: Ohio  
 COUNTRY: U.S.A.  
 ZIP: 44114-2518  
 COMPUTER READABLE FORM:  
 COMPUTER: IBM PS/2, Model 35 SX  
 OPERATING SYSTEM: DOS 5.0  
 SOFTWARE: Word Perfect 5.1  
 CURRENT APPLICATION DATA:  
 MEDIUM TYPE: Diskette, 3.50 Inch,  
 COMPUTER: IBM PS/2, Model 35 SX  
 OPERATING SYSTEM: DOS 5.0  
 SOFTWARE: Word Perfect 5.1  
 FILING DATE:  
 ATTORNEY/AGENT INFORMATION:  
 CLASSIFICATION: 435  
 NAME: Minich, Richard J  
 REGISTRATION NUMBER: 28,665  
 REFERENCE/DOCKET NUMBER: 40397/104/BABR  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (202)672-5300  
 TELEFAX: (202)672-5399  
 TELEX: 904136

INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:  
 LENGTH: 546 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear

; US-08-494-168-10  
 Query Match, Best Local Similarity 75.0%; Score 56; DB 1; Length 546;  
 Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

RESULT 24  
 US-08-494-168-10  
 Sequence 10, Application US/08494168  
 PATENT NO. 5731192  
 GENERAL INFORMATION:  
 APPLICANT: Reeders, Stephen T.  
 APPLICANT: Zhou, Jing  
 TITLE OF INVENTION: Collagen COL4A6: Gene, Protein and Method  
 TITLE OF INVENTION: of Detecting Collagen Deficiency  
 NUMBER OF SEQUENCES: 10  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Foley & Lardner  
 STREET: 3000 K Street, N.W., Suite 500  
 CITY: Washington, D.C.  
 COUNTRY: USA  
 ZIP: 20007-5109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patientin Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/494,168  
 FILING DATE:

CLASSIFICATION: 435  
 PRIORITY APPLICATION DATA:

APPLICATION NUMBER: US 08/112,465  
 FILING DATE: 27-AUG-1993

ATTORNEY/AGENT INFORMATION:  
 NAME: Saxe, Bernhard D.

REGISTRATION NUMBER: 28,665  
 REFERENCE/DOCKET NUMBER: 40397/104/BABR  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (202)672-5300  
 TELEFAX: (202)672-5399  
 TELEX: 904136

INFORMATION FOR SEQ ID NO: 15:

SEQUENCE CHARACTERISTICS:  
 LENGTH: 546 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear

; US-08-494-168-10  
 Query Match, Best Local Similarity 75.0%; Score 56; DB 1; Length 546;  
 Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

RESULT 25  
 US-09-011-735-2  
 Sequence 2, Application US/09011735B  
 PATENT NO. 6110708  
 GENERAL INFORMATION:  
 APPLICANT: Wakamatsu, No. 6110708utaka  
 TITLE OF INVENTION: Recombinant Conglutinin and Producing Method Thereof  
 FILE REFERENCE: 19056/3MS48  
 CURRENT APPLICATION NUMBER: US/09/011,735B  
 CURRENT FILING DATE: 1998-05-22  
 EARLIER APPLICATION NUMBER: JP 7-209698  
 EARLIER FILING DATE: 1995-08-17  
 NUMBER OF SEQ ID NOS: 6  
 SOFTWARE: Patentin Ver. 2.0  
 SEQ ID NO. 2  
 LENGTH: 171

; US-08-893-467A-2  
 Query Match, Best Local Similarity 60.9%; Score 56; DB 3; Length 518;  
 Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

RESULT 23  
 US-08-893-467A-2  
 Sequence 2, Application US/08893467A

PATENT NO. 6063901  
 GENERAL INFORMATION:  
 APPLICANT: Tryggason, Karl  
 APPLICANT: Elomaa, Outi  
 APPLICANT: Kangas, Marit  
 TITLE OF INVENTION: An Insolated DNA Sequence For a  
 Patent No. 6063901  
 TITLE OF INVENTION: NO. 6063901el Macrophage Receptor with  
 a Collagenous Domain and the  
 Polypeptide Chain Encoded by  
 NUMBER OF SEQUENCES: 2

CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Fay Sharpe, Beall, Fagan,  
 STREET: 1100 Superior Avenue  
 CITY: Cleveland  
 STATE: Ohio  
 COUNTRY: U.S.A.  
 ZIP: 44114-2518  
 COMPUTER READABLE FORM:  
 COMPUTER: IBM PS/2, Model 35 SX  
 OPERATING SYSTEM: DOS 5.0  
 SOFTWARE: Word Perfect 5.1  
 CURRENT APPLICATION DATA:  
 MEDIUM TYPE: Diskette, 3.50 Inch,  
 COMPUTER: IBM PS/2, Model 35 SX  
 OPERATING SYSTEM: DOS 5.0  
 SOFTWARE: Word Perfect 5.1  
 FILING DATE:  
 ATTORNEY/AGENT INFORMATION:  
 CLASSIFICATION: 435  
 NAME: Minich, Richard J  
 REGISTRATION NUMBER: 28,665  
 REFERENCE/DOCKET NUMBER: 40397/104/BABR  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (202)672-5300  
 TELEFAX: (202)672-5399  
 TELEX: 904136

INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:  
 LENGTH: 546 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear

; US-08-494-168-10  
 Query Match, Best Local Similarity 75.0%; Score 56; DB 1; Length 546;  
 Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

TYPE: PRT  
; ORGANISM: Bovine  
; US-09-011-735-2

Query Match 59.8%; Score 55; DB 3; Length 171;  
; Best Local Similarity 66.7%; Pred. No. 0.53;  
; Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
; PATENT NO. 6110708

GENERAL INFORMATION:  
; APPLICANT: Wakamiya, No. 6110708utaka  
; TITLE OF INVENTION: Recombinant Conglutinin and Producing Method Thereof  
; FILE REFERENCE: 19036/34548  
; CURRENT APPLICATION NUMBER: US/09/011,735B  
; CURRENT FILING DATE: 1998-05-22  
; EARLIER FILING DATE: JP 7-209698  
; NUMBER OF SEQ ID NOS: 6  
; SEQ ID NO 1  
; LENGTH: 351  
; SOFTWARE: PatentIn Ver. 2.0  
; TYPE: PRT  
; ORGANISM: Bovine  
; OS-09-011-735-1

RESULT 26  
US-09-011-735-1  
; Sequence 1, Application US/09/011,735B  
; PATENT NO. 6110708

GENERAL INFORMATION:  
; APPLICANT: Wakamiya, No. 6110708utaka  
; TITLE OF INVENTION: Recombinant Conglutinin and Producing Method Thereof  
; FILE REFERENCE: 19036/34548  
; CURRENT APPLICATION NUMBER: US/09/011,735B  
; CURRENT FILING DATE: 1998-05-22  
; EARLIER FILING DATE: JP 7-209698  
; NUMBER OF SEQ ID NOS: 6  
; SEQ ID NO 1  
; LENGTH: 351  
; SOFTWARE: PatentIn Ver. 2.0  
; TYPE: PRT  
; ORGANISM: Bovine  
; OS-09-011-735-1

RESULT 27  
US-08-494-168-2  
; Sequence 2, Application US/08/494,168  
; PATENT NO. 5731192

GENERAL INFORMATION:  
; APPLICANT: Reeders, Stephen T.  
; APPLICANT: Zhou, Jing  
; TITLE OF INVENTION: Gene, Protein and Method  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 3000 K Street, N.W., Suite 500  
; CITY: Washington, D.C.  
; COUNTRY: USA  
; ZIP: 20007-5109

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/494,168  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/112,465  
; FILING DATE: 27-AUG-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Saxe, Bernhard D.  
; REGISTRATION NUMBER: 28,665

RESULT 28  
US-09-219-849-38  
; Sequence 38, Application US/09/219,849  
; PATENT NO. 6150081

GENERAL INFORMATION:  
; APPLICANT: VAN HEERDE, GEORGE V.  
; APPLICANT: VAN RIJN, ALEXIS C.  
; APPLICANT: BOUMSTRA, JAN B.  
; APPLICANT: DE WOLF, FREDERIK A.  
; APPLICANT: MOOBROEK, ANDREAS  
; APPLICANT: WIND, RICHELE D.  
; APPLICANT: VAN DEN BOSCH, TANJA J.  
; TITLE OF INVENTION: SUITABLE FOR PHOTOGRAPHIC APPLICATION AND ALSO THE  
; FILE REFERENCE: 2728-2  
; CURRENT APPLICATION NUMBER: US/09/219,849  
; CURRENT FILING DATE: 1998-12-23  
; NUMBER OF SEQ ID NOS: 50  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 38  
; LENGTH: 228  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence  
; OTHER INFORMATION: amino acid sequence  
; OS-09-219-849-38

RESULT 29  
US-08-794-795-2  
; Sequence 2, Application US/08/794,795  
; PATENT NO. 5916766

GENERAL INFORMATION:  
; APPLICANT: Eishourilagay, Nabil  
; APPLICANT: Adamou, John  
; APPLICANT: Gross, Mitchell  
; APPLICANT: Lysko, Paul  
; TITLE OF INVENTION: Human Macro Scavenger Rec  
; NUMBER OF SEQUENCES: 9

REFERENCE/DOCKET NUMBER: 40397/104/BABR  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 672-5300  
; TELEFAX: (202) 672-5399  
; TELEX: 904136  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1694 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-494-168-2

Query Match 59.8%; Score 55; DB 1; Length 1694;  
; Best Local Similarity 60.0%; Pred. No. 0.99;  
; Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
; PATENT NO. 749 GPRGRKGLPGFPGLP 763

CORRESPONDENCE ADDRESS:  
 ADDRESSEE: SmithKline Beecham Corporation  
 STREET: 709 Swede-Hill Road  
 CITY: King of Prussia  
 STATE: PA  
 COUNTRY: USA  
 ZIP: 19406  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette  
 COMPUTER: IBM Compatible  
 OPERATING SYSTEM: DOS  
 SOFTWARE: FASTSEQ for Windows version 2.0  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/794,795  
 FILING DATE: 04-FEB-1997  
 CLASSIFICATION: 435  
 PRIORITY APPLICATION DATA:  
 APPLICATION NUMBER: ATG50009P  
 FILING DATE: 22-MAY-1996  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Han, William T  
 REGISTRATION NUMBER: 34,344  
 REFERENCE/DOCKET NUMBER: ATG50009  
 TELEPHONE: 610-270-5219  
 TELEFAX: 610-270-4026  
 INFORMATION FOR SEQ ID NO: 2:  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 495 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 ;US-08-794-795-2

Query Match 57.6%; Score 53; DB 2; Length 495;  
 Best Local Similarity 60.0%; Pred. No. 3.1;  
 Matches 9; Conservative 1; Mismatches 5;  
 Qy 1 GVKGDKGNGPGWPGAP 15  
 | | | : | | |  
 Db 270 GAKGDQGQPGLOGVP 284

RESULT 30  
 US-09-249-200-2  
 Sequence 2, Application US/09249200  
 Patent No. 6197931  
 GENERAL INFORMATION:  
 APPLICANT: ELSHOURBACY, NABIL  
 APPLICANT: ADAMOU, JOHN  
 APPLICANT: GROSS, MITCHELL  
 APPLICANT: LYSKO, PAUL  
 TITLE OF INVENTION: HUMAN MARCO SCAVENGER RECEPTOR  
 NUMBER OF SEQUENCES: 9  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Rarner & Prestia  
 STREET: P.O. Box 980  
 CITY: Valley Forge  
 STATE: PA  
 COUNTRY: USA  
 ZIP: 19482  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette  
 COMPUTER: IBM Compatible  
 OPERATING SYSTEM: DOS  
 SOFTWARE: FASTSEQ for Windows Version 2.0  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/249,200  
 FILING DATE: 12-FEB-1999  
 CLASSIFICATION:  
 PRIORITY APPLICATION DATA:

APPLICATION NUMBER: 08/794,795  
 FILING DATE: 04-FEB-1997  
 APPLICATION NUMBER: 60/017,699  
 FILING DATE: 23-MAY-1996  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Prestia, Paul F  
 REGISTRATION NUMBER: 23,031  
 REFERENCE/DOCKET NUMBER: ATG-50009-1  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 610-407-0700  
 TELEFAX: 610-407-0700  
 INFORMATION FOR SEQ ID NO: 2:  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 495 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 ;US-09-249-200-2

Query Match 57.6%; Score 53; DB 4; Length 495;  
 Best Local Similarity 60.0%; Pred. No. 3.1;  
 Matches 9; Conservative 1; Mismatches 5;  
 Qy 1 GVKGDKGNGPGWPGAP 15  
 | | | : | | |  
 Db 270 GAKGDQGQPGLOGVP 284

Search completed: November 1, 2002, 12:54:44  
 Job time : 11.5 secs

PAGE BLANK (USP70)

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

## OM protein - protein search, using sw model.

Run on: November 1, 2002, 12:38:32 ; Search time 13.5 Seconds

(without alignments) 106.766 Million cell updates/sec

Title: US-09-529-691A-1  
Perfect score: 92  
Sequence: 1 GVKGDKGNPGWPGAP 15

Scoring table: BLOSUM62 , Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 71:  
1: pir1:  
2: pir2:  
3: pir3:  
4: pir4:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query	Match Length	DB ID	Description
1	92	100.0	1669	1 CGHU4B	CGHU4B collagen alpha 1(IV) chain precursor - human
2	78	84.8	1669	1 CGM54B	N Alternate names: procollagen alpha 1(IV) chain
3	72	78.3	356	1 S1607	C Species: Homo sapiens (man)
4	69	75.0	1691	1 S22117	C Date: 28-May-1986 #sequence-revision 31-Dec-1992 #text-change 07-Dec-1999
5	66	71.7	754	2 A52627	C Accession: S16876; A32117; S02738; S0048; S25826; A23115; S00207; S39614; A02863;
6	66	70.7	1707	2 A33526	R Sooininen, R.; Huotari, M.; Ganguly, A.; Prockop, D.J.; Tryggvason, K. J. Biol. Chem. 264, 13565-13571, 1989
7	64	69.6	1712	1 CGHU2B	A Title: Structural organization of the gene for the alpha-1 chain of human type IV collagen
8	63	68.5	677	2 S22949	A Reference number: S16876; MUID:89340433
9	63	68.5	2944	2 A56849	A Status: nucleic acid sequence not shown; translation not shown
10	62	67.4	230	2 A32249	A Molecule type: DNA
11	62	67.4	1414	1 S2309	A Residues: 1-1659 <SO11>
12	67.4	1670	1 CGHU3B	A Cross-references: EMBL:J04217; NID:9180759; PIDN:AAA53097.1; PID:g533233	
13	61	66.3	340	1 B46345	R Poeschl, E.; Polinner, R.; Kuehn, K. EMBO J. 7, 2687-2694, 1988
14	61	66.3	1603	2 S23810	A Title: The genes for the alpha1(IV) and alpha2(IV) chains of human basement membran
15	61	66.3	1690	1 CGHU1B	A Reference number: S02738; MUID:89030632
16	61	66.3	1763	2 S13366	A Status: translation not shown
17	60	65.2	310	2 150696	A Molecule type: protein
18	60	65.2	623	2 A45337	A Residues: 1-66 <POE>
19	60	65.2	636	2 S41067	R Brazel, D.; Oberbaumer, I.; Dierlinger, H.; Babel, W.; Glanville, R.W.; Deutzhmann, Eur. J. Biochem. 168, 529-536, 1987
20	60	65.2	886	2 150694	A Title: Completion of the amino acid sequence of the alpha1 chain of human basement membran
21	60	65.2	1049	1 CGBOTS	A Reference number: S00048; MUID:88029471
22	60	65.2	1465	1 CGH07L	A Accession: S00048
23	60	65.2	1532	2 A61262	A Molecule type: mRNA
24	60	65.2	1761	2 T13990	A Residues: 1-318; 'A', 320-944 <BRAl>
25	59.5	64.7	1752	1 A44407	A Cross-references: EMBL:X05561; NID:930066; PIDN:CAA29075.1; PID:g30067
26	59	64.1	359	2 S33603	A Accession: S25826
27	59	64.1	471	2 A39024	A Molecule type: protein
28	59	64.1	1433	2 A46053	A Residues: 1-318; 'A', 320-554 <BRAl>
29	64.1	64.1	2 1744	2 S40991	R Glanville, R.W.; Olan, R.O.; Siebold, B.; Ristelli, J.; Kuehn, K. Eur. J. Biochem. 152, 213-219, 1985

## ALIGNMENTS

30	58	63.0	285	2 S60598	collagen alpha 1(I) pro
31	58	63.0	464	2 S59513	hypothetical prote
32	58	63.0	1758	2 T29351	collagen alpha 2(I)
33	58	63.0	1759	2 JC1448	collagen col-34 -
34	57	62.0	298	2 T22002	hypothetical calc pro
35	57	62.0	650	2 A45748	collagen alpha 1(V)
36	57	62.0	920	2 148103	type VII collagen
37	57	62.0	1349	2 T2833	hypothetical prote
38	56	60.9	299	2 E8633	protein F5633.1
39	56	60.9	371	2 17338	macrophage bacteri
40	56	60.9	380	2 A48295	collagen alpha 2(I)
41	56	60.9	518	2 A52840	collagen alpha 1(X)
42	56	60.9	775	2 A61228	collagen alpha 1(X)
43	56	60.9	1142	2 JX0369	macrophage scaveng
44	55	60.3	350	2 173369	macrophage scaveng
45	55.5	60.3	458	2 B44407	macrophage scaveng

A:Experimental source: Placenta	C:Complex: type IV collagen is a heterotrimer of two alpha1(IIV) chains and one alpha2(IIV) chain
A:Note: the amino end of the mature form is blocked	C:Associations among trimer amino-terminal domains (disulfide and desmosine cross-links), r-trimer associations in the interrupted helical domain (with disulfide and desmosine)
R:Soininen, H.; Ranta, R.; Hakkarainen, T.; Prockop, D.J.; Tryggvason, K.	C:Function:
FBS Lett. 225, 188-194, 1987.	A:Title: Complete primary structure of the alpha(1)-chain of human basement membrane (tyrosine-rich)
Eur. J. Biochem. 12, 4795-4802, 1993	A:Reference number: S00207; MUID:88083584
R:Babel, W.; Glanville, R.W.	A:Accession: S00207
A:Molecule type: protein	A:Molecule type: mRNA
A:Residues: 371-554 <BBL>	A:Residues: 244-530 <SOI3>
Eur. J. Biochem. 143, 545-556, 1984	A:Cross-references: EMBL:Y00706; NID:929548; PIDN:CAA68698.1; PID:929549
A:Title: Structure of human basement-membrane (type IV) collagen. Complete amino-acid sequence	R:Reble, J.A.; Golbik, R.; Mann, K.; Kuehn, K.
A:Reference number: A02863; MUID:85003629	EMBO J. 12, 4795-4802, 1993
A:Accession: A02863	A:Title: The alpha-1(beta-1 integrin recognition site of the basement membrane collagen
A:Molecule type: protein	A:Reference number: S39614; MUID:94038965
A:Residues: 371-554	A:Accession: S39614
R:Glanville, R.W.	A:Cross-references: EMBL:Y00706; NID:929548; PIDN:CAA68698.1; PID:929549
A:Title: Pepsin fragments of human placental basement-membrane collagens showing interruption	R:Babel, W.; Glanville, R.W.
A:Reference number: S16908; MUID:82005835	A:Title: Structure of human basement-membrane (type IV) collagen. Complete amino-acid sequence
A:Accession: A58317	A:Reference number: A02863; MUID:85003629
A:Molecule type: protein	A:Accession: A02863
A:Residues: 534-537, 'G', 539, 'G', 541-542, 'X', 544-553, 1389-1405, 'XX', 1408-1409, 'X', 1411-14	A:Molecule type: protein
R:Mac Wright, R.S.; Benson, V.A.; Lovello, K.T.; van der Rest, M.; Fietzek, P.P.	A:Residues: 534-718, 'D', 720-836, 'Y', 838-841, 'P', 843-903, 'Q', 905-913, 'K', 915-997, 'K', 999-
Biochemistry 22, 4940-4948, 1983	A:Experimental source: placenta
A:Title: Isolation and characterization of pepsin-solubilized human basement membrane (type IV) collagen. Comparison with the corresponding bovine protein	R:Glanville, R.W.; Rauter, A.
A:Reference number: S16910; MUID:84053346	A:Title: Pepsin fragments of human placental basement-membrane collagens showing interruption
A:Accession: S16910	A:Reference number: S16908; MUID:82005835
A:Molecule type: protein	A:Accession: A58317
A:Residues: 534-537, 'G', 539, 'G', 541-542, 'X', 544-549, 939-940, 'M', 942-944, 'V', 946, 'X', 948-	A:Residues: 534-537, 'G', 539, 'G', 541-542, 'X', 544-553, 1389-1405, 'XX', 1408-1409, 'X', 1411-14
A:Experimental source: placenta	R:Mac Wright, R.S.; Benson, V.A.; Lovello, K.T.; van der Rest, M.; Fietzek, P.P.
A:Title: Isolation and characterization of pepsin-solubilized human basement membrane (type IV) collagen. Comparison with the corresponding bovine protein	A:Title: Pepsin fragments of human placental basement-membrane collagens showing interruption
A:Reference number: S16910; MUID:84053346	A:Reference number: S16908; MUID:82005835
A:Title: cDNA clones coding for the Pro-alpha-1(IV) chain of human type IV procollagen I	A:Title: Pepsin fragments of human placental basement-membrane collagens showing interruption
A:Reference number: S01466; MUID:85207819	A:Reference number: S16908; MUID:82005835
A:Accession: S01466	A:Accession: S16908
A:Molecule type: mRNA	A:Accession: S16908
A:Residues: 1256-1669 <PIH>	A:Cross-references: EMBL:Y00940; NID:9180421; PIDN:AAA52006.1; PID:9180424
A:Cross-references: EMBL:Y00940; NID:9180421; PIDN:AAA52006.1; PID:9180424	R:Binkley, J.M.; Gudas, L.J.; Ioidl, H.R.; Wang, S.Y.; Rosenblum, J.; Kefalides, N.A.; Proc. Natl. Acad. Sci. U.S.A. 82, 3649-3653, 1985
A:Title: Restricted homology analysis of human alpha-1 type IV and other procollagen chains.	J. Biol. Chem. 260, 7687-7691, 1985
A:Reference number: S16879; MUID:85216555	A:Title: cDNA clones coding for the Pro-alpha-1(IV) chain of human type IV procollagen I
A:Accession: S16879	A:Reference number: S01466; MUID:85207819
A:Molecule type: mRNA	A:Accession: S01466
A:Residues: 1259-1669 <BRI>	A:Cross-references: EMBL:Y00940; NID:9180421; PIDN:AAA52006.1; PID:9180424
A:Cross-references: EMBL:Y00940; NID:9180421; PIDN:AAA52006.1; PID:9180424	R:Binkley, J.M.; Gudas, L.J.; Ioidl, H.R.; Wang, S.Y.; Rosenblum, J.; Kefalides, N.A.; Proc. Natl. Acad. Sci. U.S.A. 82, 3649-3653, 1985
A:Title: Restricted homology analysis of human alpha-1 type IV and other procollagen chains.	J. Biol. Chem. 260, 7687-7691, 1985
A:Reference number: S16879; MUID:85216555	A:Title: cDNA clones coding for the Pro-alpha-1(IV) chain of human type IV procollagen I
A:Accession: S16879	A:Reference number: S01466; MUID:85207819
RESULT 2	Query Match Best Local Similarity 100.0%; Score 92; DB 1; Length 1669; Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CGMS4B	QY 1 GWKGDKGNPGWGP 15
collagen alpha 1(IV) chain precursor - mouse	Dp 1263 GVKGDKGNPGWGP 1277

A; Residues: 1-185, 'L', 187-318, 'S', 320-368, 'L', 370-402, 'F', 404-480, 'L', 482-492, 'H', 494-71  
A; Cross-references: EMBL:06777  
R; Killen, P.D.; Burboeuf, P.; Sakurai, Y.; Yamada, Y.  
J. Biol. Chem. 263, 8706-8709, 1988

A; Title: structure of the amino-terminal portion of the murine alpha-1(IV) collagen chain  
A; Reference number: A28066; MUID:88243724  
A; Accession: A28066  
A; Molecule type: mRNA  
A; Residues: 1-129 <K1>  
A; Cross-references: EMBL:J03758; NID:9192669; PIDN:AAA37439.1; PID:9192670  
R; Oberbaumer, I.; Laurent, M.; Schwarz, U.; Sakurai, Y.; Yamada, Y.; Vogeli, G.; Voss, Eur. J. Biochem. 147, 217-224, 1985

A; Title: Amino acid sequence of the non-collagenous globular domain (NC1) of the alpha-1  
A; Reference number: A02864; MUID:85127033  
A; Accession: A02864  
A; Molecule type: mRNA  
A; Residues: 1276-1669 <OBE>  
A; Cross-references: EMBL:X02201; NID:950233; PIDN:CAA26132.1; PID:91333876  
R; Nath, P.; Laurent, M.; Horn, E.; Sobel, M.E.; Zon, G.; Vogeli, G.  
Gene 43, 301-304, 1986

A; Title: Isolation of an alpha-1 type-IV collagen cDNA clone using a synthetic oligodeox  
A; Reference number: A25636; MUID:86501886  
A; Accession: A25636  
A; Molecule type: mRNA  
A; Residues: 1149-1396, 'S', 1398-1424 <NAT>  
A; Cross-references: EMBL:MA4042; NID:9192286; PIDN:AAA37342.1; PID:9192287  
A; Note: the authors translated the codon CAG for residue 1374 as Arg  
R; Kurkinen, M.; Condon, M.R.; Blumberg, B.; Barlow, D.P.; Quinones, S.; Saus, J.; Pihlaj  
J. Biol. Chem. 262, 8496-8499, 1987

A; Title: Extensive homology between the carboxyl-terminal peptides of mouse alpha-1(IV)  
A; Reference number: A94680; MUID:87250460  
A; Accession: A29501  
A; Molecule type: mRNA  
A; Residues: 141-1669 <KUR>  
A; Cross-references: EMBL:MI5832; NID:9192288; PIDN:AAA37340.1; PID:9387115  
R; Killen, P.D.; Burboeuf, P.D.; Martin, G.R.; Yamada, Y.  
J. Biol. Chem. 263, 12310-12314, 1988

A; Title: Characterization of the promoter for the alpha-1(IV) collagen gene. DNA sequenc  
A; Reference number: S19079; MUID:88315019  
A; Accession: S19079  
A; Molecule type: DNA  
A; Residues: 1-28 <KIT>  
A; Cross-references: EMBL:J03944; NID:9192673; PIDN:AAA37442.1; PID:9466503  
R; Kaytes, P.; Wood, L.; Theriault, N.; Kurkinen, M.; Vogeli, G.  
J. Biol. Chem. 263, 19274-19277, 1988

A; Title: Head-to-head arrangement of murine type IV collagen genes.  
A; Reference number: A92702; MUID:89066738  
A; Accession: A32003  
A; Molecule type: DNA  
A; Residues: 1-28 <KAY>  
A; Cross-references: EMBL:J04448; NID:9192666; PIDN:AAA37437.1; PID:9450449  
R; Burboeuf, P.D.; Martin, G.R.; Yamada, Y.  
Proc. Natl. Acad. Sci. U.S.A. 85, 9679-9682, 1988

A; Title: Alpha1(IV) and alpha2(IV) collagen genes are regulated by a bidirectional prom  
A; Reference number: A94220; MUID:89071759  
A; Accession: A31766  
A; Molecule type: DNA  
A; Residues: 1-28 <BUR>  
A; Cross-references: EMBL:M23333; NID:9340878; PIDN:AAA51625.1; PID:9535668  
R; Sakurai, Y.; Sullivan, M.; Yamada, Y.  
J. Biol. Chem. 261, 6654-6657, 1986

A; Title: Alpha-1 type IV collagen gene evolved differently from fibrillar collagen genes  
A; Reference number: S19094; MUID:86196099  
A; Molecule type: DNA  
A; Residues: 110-1135,1189-1316,1342-1383,1418-1487 <SAK>  
A; Cross-references: EMBL:MI1307  
R; Schuppan, D.; Timpl, R.; Glanville, R.W.  
EBS Lett. 115, 297-300, 1980

A; Title: Discontinuities in the triple helical sequence Gly-X-Y of basement membrane (ty  
A; Reference number: S16909; MUID:80246483  
A; Accession: S16909  
A; Molecule type: protein

A; Residues: 940-946, 'G', 948-949, 'G', 951-955, 'G', 957-1213-1228, 'X', 1230-1234, 'P', 1236-  
A; Cross-references: EMBL:06777  
R; Schuppan, D.; Glanville, R.W.; Timpl, R.  
R; Killen, P.D.; Burboeuf, P.; Sakurai, Y.; Yamada, Y.  
J. Biol. Chem. 263, 505-512, 1982

A; Title: Covalent structure of mouse type-IV collagen. Isolation, order and partial a  
A; Reference number: A25991; MUID:82186723  
A; Accession: A25991  
A; Molecule type: protein  
A; Residues: 940-946, 'X', 948-949, 'X', 951-955, 'X', 957-964, 'X', 966-991, 'X', 993-1003, 'X'  
A; Accession: B25991  
A; Molecule type: protein  
A; Residues: 1173-1181, 'X', 1183-1184, 'X', 1185-1187, 'X', 1189-1205, 'O', 1207, 'X', 1210-12  
A; Accession: B25991  
A; Molecule type: protein  
A; Residues: 1173-1181, 'X', 1183-1184, 'X', 1185-1187, 'X', 1189-1205, 'O', 1207, 'X', 1210-12  
3, 'SP', 1266, 'IT', 1269, 'SK', 1272, 'DM', 1275, 'L', 1277-1282, 1316-1318, 'X', 1320-1327, 'X', 1  
R; Weber, S.; Engel, J.; Wiedemann, H.; Glanville, R.W.; Timpl, R.  
R; J. Biochem. 139, 401-410, 1984

A; Title: Subunit structure and assembly of the globular domain of basement-membrane c  
A; Reference number: S17801; MUID:84132058  
A; Accession: S17801  
A; Molecule type: protein  
A; Residues: 1435-1443 <WEB>  
A; Genetics:  
A; Introns: 28/3; 48/3; 78/3; 93/3; 108/3; 129/3  
A; Note: the list of introns may be incomplete  
C; Superfamily: collagen alpha 1(IV) chain  
C; Keywords: basement membrane; cell binding; coiled coil; duplication; extracellular  
F; 1-27/Domain: signal sequence #status predicted <SIG>  
F; 28-169/Domain: 7S <7SD>  
F; 163-1440/Domain: collagenous, triple helix <COL>  
F; 597-599/Region: cell attachment (R-G-D) motif  
F; 781-783/Region: cell attachment (R-G-D) motif  
F; 917-919/Region: cell attachment (R-G-D) motif  
F; 968-970/Region: cell attachment (R-G-D) motif  
F; 141-169/Region: duplication  
F; 1441-1552/Region: duplication  
F; 31-36, 39, 41, 434, 467, 470/Disulfide bonds: interchain #status predicted  
F; 126/Domain: binding site: carbohydrate (Asn) (covalent) #status predicted  
F; 971-974, 977, 986, 989, 1001, 1007, 1017, 1019, 1022, 1031, 1037, 1040, 1055, 1060, 1063, 1075, 1078, 109  
92, 1298, 1310, 1313, 1322, 1337, 1346, 1349, 1422, 1425, 1431, 1437, 1440/Modified site: hydroxy  
F; 1214-1424/Modified site: 4-hydroxyproline (Pro) #status experimental  
F; 1304/Modified site: 5-hydroxylysine (Lys) #status experimental  
F; 1505-1511, 1616-1622/Disulfide bonds: #status predicted  
F; 31-36, 39, 41, 434, 467, 470/Disulfide bonds: interchain #status predicted  
F; 126/Domain: binding site: carbohydrate (Asn) (covalent) #status predicted  
F; 971-974, 977, 986, 989, 1001, 1007, 1017, 1019, 1022, 1031, 1037, 1040, 1055, 1060, 1063, 1075, 1078, 109  
92, 1298, 1310, 1313, 1322, 1337, 1346, 1349, 1422, 1425, 1431, 1437, 1440/Modified site: hydroxy  
F; 1214-1424/Modified site: 4-hydroxyproline (Pro) #status experimental  
F; 1505-1511, 1616-1622/Disulfide bonds: #status predicted  
Query Match Best Local Similarity Score 84.8%; Length 1669;  
Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Qy 1 GVKGDKGNGPGRGP 15  
Db 1263 GPKGDKGNQPGAP 1277

RESULT 3  
S16907  
collagen alpha 1(IV) chain - bovine (fragments)  
C; Species: Bos primigenius tauris (cattle)  
C; Date: 04-Dec-1992 #sequence\_revision 03-Apr-1995 #text\_change 19-Oct-1995  
C; Accession: A39474; S16907; S58432; J39319  
R; Gunwani, S.; Noebl, B.G.  
J. Biol. Chem. 266, 14088-14094, 1991  
A; Title: Properties of the collagenous domain of the alpha3(IV) chain, the goodpastur  
A; Reference number: A39474; MUID:9310700  
A; Accession: A39474  
A; Molecule type: protein  
A; Residues: 1-9 <GU2>  
R; Schuppan, D.; Glanville, R.W.; Timpl, R.; Dixit, S.N.; Kang, A.H.  
J. Biol. Chem. 220, 227-233, 1984  
A; Title: Sequence comparison of pepsin-resistant segments of basement-membrane colla  
A; Reference number: S16907; MUID:84256630  
A; Accession: S16907  
A; Molecule type: protein

A; Residues: 4-8; 'E', 110-68-69-150; 150-322-322-349	R; Bortnowski, R.J.; Langfeld, J.P.M.; Wieslander, J.; Hamilton, J.; Hudson, B.G.	J. Biol. Chem. 262, 7874-7877, 1987
A; Title: Localization of the Goodpasture epitope to a novel chain of basement membrane		
A; Reference number: S18432; MUID:87222419		
A; Accession: S18432		
A; Molecule type: protein		
A; Residues: 337-347, 'R', 350-353 <BU2>		
A; Residues: 337-347, 'FL', 350-356 <GU3>		
C; Superfamily: collagen alpha 1(IV) chain		
C; Keywords: basement membrane; cell binding; coiled coil; disulfide bond; extracellular		
F; 31; 34, 37-46, 61, 69-78, 84, 87, 102, 110, 122, 125, 137, 140, 143, 149, 155, 158, 161, 164, 185, 188, 199		
A; Title: Glomerular basement membrane. Identification of dimeric subunits of the noncollagenous (Pro) #status experimental		
A; Reference number: A39419; MUID:91332055		
A; Accession: A39419		
A; Molecule type: protein		
A; Residues: 337-347, 'FL', 350-356 <GU3>		
A; Best Local Similarity 66.7%; Pred. No. 0; 0.003; Matches 10; conservative 3; Mismatches 2; Indels 0; Gaps 0;		
Qy 1 GVKGDKGNPGRGAP 15		
Db 248 GLXGERGNPGWPCTP 262		
RESULT 4		
S22917		
A; Molecule type: collagen alpha 5(IV) chain precursor, renal splice form - human		
A; Residues: 1-967 <ZHO>		
A; Cross-references: GB:M90464; NID:9180826; PIDN:AAA52046.1; PID:9180824		
R; Zhou, J.J.; Leinonen, A.; Tryggvason, K.		
J. Biol. Chem. 267, 12475-12481, 1992		
A; Title: Complete amino acid sequence of the human alpha-5(IV) collagen chain and identifi		
n Alport syndrome patient.		
A; Reference number: S22917; MUID:92316923		
A; Accession: S22917		
A; Molecule type: mRNA		
A; Residues: 1-967 <ZHO>		
A; Cross-references: GB:M90464; NID:9180826; PIDN:AAA52046.1; PID:9180824		
R; Zhou, J.J.; Leinonen, A.; Tryggvason, K.		
J. Biol. Chem. 269, 6608-6614, 1994		
A; Title: Structure of the human type IV collagen COL4A5 gene.		
A; Reference number: A54365; MUID:94165049		
A; Accession: A54365		
A; Molecule type: DNA		
A; Residues: 1-922 <ZH2>		
A; Cross-references: GB:U04470; NID:9463378; GB:U04520; NID:9463428; PIDN:ACC27816.1; PID:9463428		
R; Zhou, J.; Mochizuki, T.; Smets, H.; Antignac, C.; Laurila, P.; de Paepe, A.; Tryggvason, K.		
Science 261, 1167-1169, 1993		
A; Title: Deletion of the paired alpha5(IV) and alpha6(IV) collagen genes in inherited syn		
A; Reference number: A57079; MUID:93361972		
A; Accession: A57079		
A; Molecule type: DNA		
A; Residues: 1-27 <ZH4>		
A; Cross-references: GB:Z37153; NID:9587203; PIDN:CA85512.1; PID:9587204		
R; Pihlajaniemi, T.; Pohjolainen, E.R.; Myers, J.C.		
J. Biol. Chem. 265, 13758-13766, 1990		
A; Title: Complete primary structure of the triple-helical region and the carboxyl-terminal		
A; Reference number: A37122; MUID:9033790		
A; Accession: A37122		
A; Molecule type: mRNA		
A; Residues: 84-439; 'GS', 442-624, 'LAQ', 629-666, 'FR', 669-887, 'R', 889-1264, 1271-1691 <PIH>		
A; Cross-references: GB:J05558; EMBL:N58526; NID:91314209		
A; Note: submitted to the EMBL data library, February 1991		
A; Note: the authors translated the codon GCC for residue 115 as Val		
R; Renieri, A.; Seri, M.; Myers, J.C.; Pihlajaniemi, T.; Massella, L.; Rizzoni, G.; De Ma		
A; Gene: GDB:COL4A5; ATC		
A; Residues: 1-1607, 'VHAYKC' <LEM>		
A; Cross-references: GB:S67677; NID:9425553; PIDN:AA13967.1; PID:94261667		
A; Note: frameshift mutation from a patient with Alport syndrome; five other mutations		
C; Comment: Prolines and lysines at the third position of the tripeptide repeating unit		
and subsequently o-glycosylated.		
A; Genetics:		

A,Cross-references: GDB:120596; OMIM:303630  
A,Map position: Xq22-Xq22  
A,Introns: 27/3; 47/3; 77/3; 92/3; 107/3; 128/3; 146/3; 155/3; 182/3; 203/3; 215/3; 229/3; 79/1; 837/1; 893/1; 923/1; 973/1; 1006/1; 1036/1; 1082/3; 1125/1; 1152/1; 1185/1; 1  
C;Note: the alpha 5(IV) and alpha 6(IV) chain genes are encoded on opposite strands with  
C:Complex: this minor type IV collagen is thought to form a heterotrimer of two alpha 5(IV)  
C:Function: trimers (with disulfide and desmosine cross-links), dimeric and  
C:Keywords: Airport syndrome; basement membrane; collagen alpha 1(IV) chain  
C:Superfamily: collagen alpha 1(IV) chain  
C:Species: Mus musculus (house mouse)  
C;Date: 04-Dec-1992 #sequence\_revision 04-Dec-1992 #text\_change 13-Aug-1999  
C;Accession: A33526; A24432; D24432; A25066; B25066; A23564; S19081; B29301; A24628; B32  
R;Saus, J.; Quinones, S.; Mackrill, A.; Blumberg, B.; Muthukumaran, G.; Pihlajaniemi, T.  
J. Biol. Chem. 264, 618-624, 1989  
A;Title: The complete primary structure of mouse alpha-2(IV) collagen. Alignment with mo  
A;Reference number: A33526; MUID:89197933  
A;Accession: A33526

**RESULT 6**

A33526  
collagen alpha 2(IV) chain precursor - mouse  
C;Species: Mus musculus (house mouse)  
C;Date: 04-Dec-1992 #sequence\_revision 04-Dec-1992 #text\_change 13-Aug-1999  
C;Accession: A33526; A24432; D24432; A25066; B25066; A23564; S19081; B29301; A24628; B32  
R;Turbel, P.D.; Martin, G.R.; Yamada, Y.  
Proc. Natl. Acad. Sci. U.S.A. 85, 6679-6682, 1988  
A;Title: Alpha1(IV) and alpha2(IV) collagen genes are regulated by a bidirectional pr  
A;Reference number: A33526; MUID:89197933  
A;Accession: A33526

A,Cross-references: EMBL:J04469; NID:9556298; PIDN:AAA50293.1; PID:9556299  
A,Residues: 1-1707 <SAU>  
A,Cross-references: EMBL:J04695; NID:9556298; PIDN:AAA50293.1; PID:9556299  
R;Kurkinen, M.; Bernard, M.P.; Barlow, D.P.; Chow, L.T.  
Nature 317, 177-179, 1985  
A;Title: Characterization of 54-, 123- and 182-base-pair exons in the mouse alpha-2(I  
A;Reference number: A33367; MUID:85296379  
A;Accession: A24432

A,Molecule type: mRNA  
A,Residues: 967-1096; 'G', 1098-1109 <KUI>  
A,Cross-references: EMBL:X02896; NID:950263; PIDN:CAA26555.1; PID:9899326  
A,Residues: 1-176 <SIG>  
A,Accession: D24432  
A,Molecule type: DNA  
A,Residues: 964-1096; 'G', 1098-1109 <KU2>  
A,Cross-references: EMBL:X02899  
R;Schwarz, U.; Schuppan, D.; Oberbaeumer, I.; Glanville, R.W.; Deutzmann, R.; Impl,  
R.; J. Biochem. 157, 49-55, 1985  
A;Title: Structure of mouse type IV collagen. Amino-acid sequence of the C-terminal 5  
A,Reference number: A25066; MUID:86220192  
A,Accession: A25066  
A,Molecule type: mRNA  
A,Residues: 970-1480 <SC1>  
A,Cross-references: EMBL:X04647  
A,Accession: B25066  
A,Molecule type: protein  
A,Residues: 979-1055-1101-1105-1222-1226-1310-1326-1335-1351-1480 <SC2>  
R;Vogell, G.; Horn, E.; Carter, J.; Kaytes, P.S.  
FEBS Lett. 206, 29-32, 1986  
A;Title: Proposed alignment of helical interruptions in the two subunits of the basem  
A,Reference number: A24364; MUID:87005245  
A,Accession: A24364  
A,Molecule type: mRNA  
A,Residues: 1041-1050; 'R', 1052-1170; 'S', 1172-1178; 'R', 1180-1240; 'E', 1242-1327; 'A', 132  
A,Cross-references: EMBL:X04410; NID:950240; PIDN:CA27998.1; PID:929678  
R;Kaytes, P.S.; Theriault, N.Y.; Vogell, G.  
Gene 54, 141-146, 1987  
A;Title: Homologies between the non-collagenous C-terminal (NC1) globular domains of  
A,Reference number: S19080; MUID:87277427  
A,Accession: S19080  
A,Molecule type: mRNA  
A,Residues: 1466-1622; 'H', 1624-1707 <KA1>  
A,Cross-references: GB:X04410; NID:950240; PIDN:CA27998.1; PID:929678  
R;Kurkinen, M.; Condron, M.R.; Blumberg, B.; Barlow, D.P.; Quinones, S.; Saus, J.; Pih  
FEMS Lett. 202, 849-852, 1987  
A;Title: Extensive homology between the carboxyl-terminal peptides of mouse alpha-1(I  
A;Reference number: A34680; MUID:87250460  
A,Accession: B29301  
A,Molecule type: mRNA  
A,Residues: 1481-1707 <KUR>  
A,Cross-references: EMBL:ML5833; NID:9192284; PIDN:AAA37341.1; PID:9387116  
R;Schwarz-Maggolen, U.; Oberbaeumer, I.; Kuehn, K.  
FEBS Lett. 208, 203-207, 1986  
A;Title: cDNA and protein sequence of the NC1 domain of the alpha-2-chain of collagen  
A,Reference number: A34680; MUID:87054581  
A,Accession: A24628  
A,Molecule type: DNA  
A,Residues: 1480-1572; 'L', 1574-1622; 'H', 1624-1707 <SCH>  
A,Cross-references: EMBL:X04647  
R;Kaytes, P.; Wood, L.; Theriault, N.; Kurkinen, M.; Vogell, G.  
J. Biol. Chem. 263, 19274-19277, 1988  
A;Title: Head-to-head arrangement of murine type IV collagen genes.  
A,Reference number: A92702; MUID:89066738  
A,Accession: B32003  
A,Molecule type: DNA  
A,Residues: 1-33 <KA2>  
A,Cross-references: EMBL:J04448; NID:9192666; PIDN:AAA37438.1; PID:9192667  
R;Turbel, P.D.; Martin, G.R.; Yamada, Y.  
Proc. Natl. Acad. Sci. U.S.A. 85, 6679-6682, 1988  
A;Title: Alpha1(IV) and alpha2(IV) collagen genes are regulated by a bidirectional pr  
A,Reference number: A34620; MUID:89071759  
A,Accession: B31766  
A,Molecule type: DNA

A; Residues: 1-60 <BUR>  
A; Cross-references: EMBL:M23333  
R; Weber, S.; Engel, J.; Wiedemann, H.; Glanville, R.W.; Timpl, R.  
Eur. J. Biochem. 139, 401-410 1984  
A; Title: Subunit structure and assembly of the globular domain of basement-membrane collagen  
A; Reference number: S1701; MUID:88132058  
A; Accession: S19086  
A; Molecule type: protein  
A; Residues: 1475-1481, X, 1483-1487 <WEBB>  
C; Genetics:  
A; Introns: 15/2; 33/3; 963/1; 1003/3; 1064/3; 1085/3  
C; Superfamily: collagen alpha 1(IV) chain  
C; Keywords: basement membrane; cell binding; coiled coil; disulfide bond; duplication; epsilon; F1; 28/Domain: signal sequence #status predicted <SIG>  
F; 29-171/Domain: 7S #status predicted <7SD>  
F; 58-1480/Domain: collagenous #status predicted <COL>  
F; 141-143/Region: cell attachment (R-G-D) motif  
F; 360-362/Region: cell attachment (R-G-D) motif  
F; 179-781/Region: cell attachment (R-G-D) motif  
F; 884-886/Region: cell attachment (R-G-D) motif  
F; 955-967/Region: cell attachment (R-G-D) motif  
F; 125-125/Region: cell attachment (R-G-D) motif  
F; 1447-1449/Region: cell attachment (R-G-D) motif  
F; 181-181 <1589/Domain: repeat NCL #status predicted <NC1>  
F; 1590-1597/Domain: repeat NCL #status predicted <NC1>  
F; 42-47,51,53,481-483/Disulfide bonds: interchain #status predicted  
F; 188-1270/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F; 656-676-1540-1550-1653-1660/Disulfide bonds: #status predicted  
F; 985, 988, 997, 1003, 1028, 1031, 1067, 1101, 1113, 1119, 1143, 1170, 1200, 1242, 1305, 1368, 1391/Modifications  
F; 985, 988, 997, 1003, 1028, 1031, 1067, 1101, 1113, 1119, 1143, 1170, 1200, 1242, 1305, 1368, 1391/Modifications  
F; 1009, 1012, 1018, 1021, 1024, 1037, 1040, 1043, 1046, 1052, 1058, 1070, 1098, 1110, 1128, 1140, 1149, 1177, 1383, 1386, 1401, 1408, 1420, 1423, 1429, 1444, 1465, 1468, 1471, 1477/Modified site: hydroxyproline  
Query Match 70.7%; Score 65; DB 2; Length 1707;  
Best Local Similarity 66.7%; Pred. No. 0.16;  
Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
Oy 1 GVKGDKGNPGWPGP 15  
Db 1165 GIPGDKGDKGPGP 1179

RESULT 7

CGH12B

collagen alpha 2(IV) chain precursor - human

N; Alternate names: procollagen alpha 2(IV) chain

C; Species: Homo sapiens (man)

C; Date: 07-Jun-1990 #sequence\_revision 03-Oct-1995 #text\_change 22-Jun-1999

C; Accession: A3204; S0007; S02624; S00246; S17678; S16911; B32117; S16877; S00165; S39

R; Hostikka, S.L.; Tryggvason, K.

J. Biol. Chem. 263, 19418-19493, 1988

A; Title: The complete primary structure of the alpha2 chain of human type IV collagen

A; Reference number: A32024; MUID:8906769

A; Accession: S12024

A; Molecule type: mRNA

A; Residues: 1-1712 <HS01>

A; Cross-references: EMBL:J04210; EMBL:X05610; GB:M20753; NID:g29550; PIDN:CAA29098.1; PII: FEB5 Lett. 215, 281-286, 1987

R; Hostikka, S.L.; Kurkinen, M.; Tryggvason, K.

FEB5 Lett. 215, 281-286, 1987

A; Title: Nucleotide sequence coding for the human type IV collagen alpha-2 chain cDNA related region.

A; Reference number: S00007; MUID:87219158

A; Molecule type: mRNA

A; Residues: 1254-1988, W, 1400-1712 <HS02>

A; Cross-references: EMBL:J04210; EMBL:X05610; GB:M20753; NID:g29550; PIDN:CAA29098.1; PII: R; Note: 1399-W was also found

R; Hostikka, S.L.; Tryggvason, K.

FEB5 Lett. 224, 297-305, 1987

A; Title: Extensive structural differences between genes for the alpha(1) and alpha(2) chains

A; Reference number: S02624; MUID:88083553

A; Accession: S02624

A; Status: not compared with conceptual translation

A; Molecule type: DNA

A; Residues: 1347-1350; 1377-1383; 1426-1432; 1465-1471; 1529-1535; 1625-1630 <HS03>

A; Note: complete nucleotide sequence not shown

R; Brazel, D.; Pollner, R.; Oberbaumer, I.; Kuehn, K.

Eur. J. Biochem. 172, 35-42, 1988

A; Title: Human basement membrane collagen (type IV): the amino acid sequence of the alpha 1(IV) chain

A; Reference number: S00246; MUID:88151998

A; Molecule type: mRNA

A; Residues: 1-682, 'G', 684-1043 <BR>

A; Cross-references: EMBL:X05562; NID:g0075; PIDN:CAA29076.1; PID:g30076

R; Oberbaumer, I.

A; Reference number: S17678

A; Accession: S17678

A; Molecule type: mRNA

A; Residues: 1-470, 'P', 472-682, 'G', 684-1043 <COBE>

A; Cross-references: EMBL:X05562; NID:g30075; PIDN:CAA29076.1; PID:g30076

R; Poeschl, E.; Pollner, R.; Kuehn, K.

EMBO J. 7, 2687-2695, 1988

A; Title: The genes for the alpha1(IV) and alpha2(IV) chains of human basement membrane

A; Reference number: S02738; MUID:89030632

A; Accession: S16911

A; Status: translation not shown

A; Molecule type: DNA

A; Residues: 1-33 <POE>

A; Cross-references: EMBL:X12784; GB:M36963; NID:g30072; PIDN:CAA31275.1; PID:g30073

R; Soininen, R.; Huotari, M.; Hostikka, S.L.; Prockop, D.J.; Tryggvason, K.

J. Biol. Chem. 263, 17217-17220, 1988

A; Title: The structural genes for alpha1 and alpha2 chains of human type IV collagen

A; Reference number: A92690; MUID:89034231

A; Accession: B32117

A; Molecule type: DNA

A; Residues: 1-33 <SC01>

A; Cross-references: EMBL:J04217; EMBL:J05039; NID:g180759; PIDN:AAA3097.1; PID:g5532

R; Soininen, R.; Huotari, M.; Ganguly, A.; Prockop, D.J.; Tryggvason, K.

J. Biol. Chem. 264, 13565-13571, 1989

A; Title: Structural organization of the gene for the alpha-1 chain of human type IV collagen

A; Reference number: S16876; MUID:89340333

A; Accession: S16877

A; Molecule type: nucleic acid

A; Status: nucleic acid sequence not shown; translation not shown

A; Molecule type: DNA

A; Residues: 1-33 <SC01>

A; Cross-references: EMBL:J04217; NID:g180759; PIDN:AAA3097.1; EMBL:J050

R; Note: this sequence was submitted to the EMBL Data Library, October 1988

R; Siebold, B.; Qian, R.Q.; Glanville, R.W.; Hofmann, R.; Kuehn, K.

Bur. J. Biochem. 108, 559-575, 1987

A; Title: Construction of a model for the aggregation and cross-linking region (7S domain) region.

A; Reference number: S00165; MUID:88029476

A; Accession: S00165

A; Molecule type: protein

A; Residues: 37-247 <SEI1>

A; Note: the sequence from Fig. 4 is inconsistent with that from Fig. 3 in having 175-R; Eble, J.A.; Golbik, R.; Mann, K.; Kuehn, K.

EMBO J. 12, 4795-4802, 1993

A; Title: Cross-linking of the alpha 1-beta 1 integrin recognition site of the basement membrane collagen

A; Reference number: S3964; MUID:94038663

A; Accession: S39615

A; Molecule type: protein

A; Residues: 407-570 <EBI>

Biochemistry 22, 4940-4948, 1983

A; Title: Isolation and characterization of pepsin-solubilized human basement membrane

A; Reference number: S16910; MUID:84053346

A; Accession: S16912

A; Molecule type: protein

A; Residues: 490-492, X, 494-496, 675-677, 'G', 679-680, 'G', 682, 684-685, 'P' <MAC>

A; Experimental source: placenta

R; Glanville, R.W.; Rauter, A.

Hoppe-Seyler's Z. Physiol. Chem. 362, 943-951, 1981

A;Title: Pepsin fragments of human placental basement-membrane collagens showing interrupted local structure. **PLAC-BM**  
A;Reference number: S16908; MUID:82005835  
A;Accession: B58517  
A;Molecule type: protein  
A;Residues: 490-492, 'X', 494-501, 'P', 503-507; 952-957, 'X', 959-966, 'X', 968; 984-986, 'X', 988-989; 1000-1002  
R;Killen, P.D.; Francomano, C.A.; Yamada, Y.; Modi, W.S.; O'Brien, S.J.  
Hun. Genet. 77, 318-324, 1987  
A;Title: Partial structure of the human alpha-2(IV) collagen chain and chromosomal localization. **COL4A2**  
A;Reference number: S01450; MUID:8805168  
A;Accession: S01450  
A;Molecule type: mRNA  
A;Residues: 1040, 'L', 1042-1398, 'V', 1400-1418, 'W', 1420-1635, 'V', 1637-1712 <KIL>  
A;Cross-references: EMBL:M24766; NID:9537328; PIDN:AAA52043.1; PID:9537329  
A;Note: the sequence form Fig. 7 is inconsistent with that shown in Fig. 11 in having 17 residues instead of 16.  
A;Reference number: A27114; MUID:87250571  
A;Gene: GDB:COL4A2  
A;Map position: 13q34-13q34  
A;Introns: 15/2; 33/3; 1347/1; 1380/1; 1429/1; 1468/1; 1532/1; 1527/3 #status incomplete and subsequently o-glycosylated.  
A;Cross-references: GDB:119792; OMIM:120090  
A;Function:  
A;Description: structural component of basement membrane  
C;Superfamily: collagen alpha 1(IV) chain  
C;Keywords: basement membrane; cell binding; coiled coil; extracellular matrix; glycoprotein  
F;1-28/domain: signal sequence #status predicted <SIG>  
F;1-28/domain: collagen alpha 2(IV) chain #status predicted <MAT>  
F;1-29-17/2/product: collagen alpha 2(IV) chain #status predicted <MAT>  
F;1-57/domain: amino-terminal nonhelical, NHI <NHI>  
F;58-145/Region: interrupted helical  
F;362-364/Region: cell attachment (R-G-D) motif  
F;786-786/Region: cell attachment (R-G-D) motif  
F;868-870/Region: cell attachment (R-G-D) motif  
F;889-891/Region: cell attachment (R-G-D) motif  
F;90-972/Region: cell attachment (R-G-D) motif  
F;1069-1071/Region: cell attachment (R-G-D) motif  
F;1228-1230/Region: cell attachment (R-G-D) motif  
F;1486-1712/domain: carboxyl-terminal nonhelical, NCI <NCI>  
F;1495-1593/domain: collagen IV carboxyl-terminal repeat <CT1>  
F;1603-1708/domain: collagen IV carboxyl-terminal repeat <CT2>  
F;42-47, 51-53, 137-483, 485/dsulfide bonds: interchain #status predicted  
F;57, 87, 90, 102, 165, 168-225, 239-242/Binding site: carbohydrate (Lys) (covalent) #status experimental  
F;57/Modified site: 5-hydroxylysine (Lys) #status atypical  
F;63-75, 96-114, 120-123, 132-150, 159-166, 189-198, 201-213, 216, 219, 496, 499, 955, 964, 1103, 1115  
F;67, 90, 12, 165, 168, 225, 239, 242/Modified site: 5-hydroxylysine (Lys) #status experimental  
F;138/Binding site: carbohydrate (Asn) (covalent) #status experimental  
F;209/Modified site: 4-hydroxyproline (Pro) #status atypical  
F;661-681/Dsulfide bonds: #status predicted  
F;1275/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F;1504-1590, 1537-1593/dsulfide bonds: (or 1504-1593, 1537-1590) #status experimental  
F;1549-1555, 1555-1565/dsulfide bonds: #status experimental  
F;1612-1705, 1646-1708/dsulfide bonds: (or 1612-1708, 1646-1705) #status experimental  
A;Title: Duplication of type IV collagen COOH-terminal repeats and species-specific expression. **COL4A2**  
A;Reference number: A27114; MUID:87250571  
A;Molecule type: mRNA  
A;Residues: 1486-1574, 'I', 1576-1712 <NVE>  
A;Cross-references: EMBL:J02760; NID:9180425; PIDN:AAA58422.1; PID:9180426  
C;Comment: Prolines and lysines at the third position of the tripeptide repeating unit are added subsequently o-glycosylated.  
C;Genetics:  
A;Gene: GDB:COL4A2  
A;Map position: 13q34-13q34  
A;Introns: 15/2; 33/3; 1347/1; 1380/1; 1429/1; 1468/1; 1532/1; 1527/3 #status incomplete and subsequently o-glycosylated.  
C;Complex: Type IV collagen is a heterotrimer of two alpha 1(IV) chains (see PIR:CGHU1B) domains (with disulfide and desmosine cross-links), dimeric associations among trimer interrupted helical domain (with disulfide and desmosine cross-links).  
C;Function:  
A;Description: structural component of basement membrane  
C;Superfamily: collagen alpha 1(IV) chain  
C;Keywords: basement membrane; cell binding; coiled coil; extracellular matrix; glycoprotein  
F;1-28/domain: signal sequence #status predicted <SIG>  
F;1-28/domain: collagen alpha 2(IV) chain #status predicted <MAT>  
F;1-57/domain: amino-terminal nonhelical, NHI <NHI>  
F;58-145/Region: interrupted helical  
F;362-364/Region: cell attachment (R-G-D) motif  
F;786-786/Region: cell attachment (R-G-D) motif  
F;868-870/Region: cell attachment (R-G-D) motif  
F;889-891/Region: cell attachment (R-G-D) motif  
F;90-972/Region: cell attachment (R-G-D) motif  
F;1069-1071/Region: cell attachment (R-G-D) motif  
F;1228-1230/Region: cell attachment (R-G-D) motif  
F;1486-1712/Region: cell attachment (R-G-D) motif  
F;1495-1593/domain: collagen IV carboxyl-terminal repeat <CT1>  
F;1603-1708/domain: collagen IV carboxyl-terminal repeat <CT2>  
F;42-47, 51-53, 137-483, 485/dsulfide bonds: interchain #status predicted  
F;57, 87, 90, 102, 165, 168-225, 239-242/Binding site: carbohydrate (Lys) (covalent) #status experimental  
F;57/Modified site: 5-hydroxylysine (Lys) #status atypical  
F;63-75, 96-114, 120-123, 132-150, 159-166, 189-198, 201-213, 216, 219, 496, 499, 955, 964, 1103, 1115  
F;67, 90, 12, 165, 168, 225, 239, 242/Modified site: 5-hydroxylysine (Lys) #status experimental  
F;138/Binding site: carbohydrate (Asn) (covalent) #status experimental  
F;209/Modified site: 4-hydroxyproline (Pro) #status atypical  
F;661-681/Dsulfide bonds: #status predicted  
F;1275/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F;1504-1590, 1537-1593/dsulfide bonds: (or 1504-1593, 1537-1590) #status experimental  
F;1549-1555, 1555-1565/dsulfide bonds: #status experimental  
F;1612-1705, 1646-1708/dsulfide bonds: (or 1612-1708, 1646-1705) #status experimental  
A;Title: The arrangement of intra- and intermolecular disulfide bonds in the carboxyterminal region of the human alpha-2(IV) collagen chain. **COL4A2**  
A;Reference number: S02250; MUID:89005112  
A;Accession: S02250  
A;Molecule type: protein  
A;Residues: 1480-1535, 1545-1614, 1617-1662, 'H', 1664-1700, 'G', 1705-1708, 1710-1712 <SIE2>  
A;Note: the sequence form Fig. 7 is inconsistent with that shown in Fig. 11 in having 17 residues instead of 16.  
A;Cross-references: EMBL:M24766; NID:9537328; PIDN:AAA52043.1; PID:9537329  
A;Title: The arrangement of intra- and intermolecular disulfide bonds in the carboxyterminus of the human alpha-2(IV) collagen chain. **COL4A2**  
A;Reference number: S02250; MUID:89005112  
A;Accession: S02250  
A;Molecule type: mRNA  
A;Residues: 1480-1535, 1545-1614, 1617-1662, 'H', 1664-1700, 'G', 1705-1708, 1710-1712 <KIL>  
A;Title: Duplication of type IV collagen COOH-terminal repeats and species-specific expression. **COL4A2**  
A;Reference number: A27114; MUID:87250571  
A;Gene: GDB:COL4A2  
A;Map position: 13q34-13q34  
A;Introns: 15/2; 33/3; 1347/1; 1380/1; 1429/1; 1468/1; 1532/1; 1527/3 #status incomplete and subsequently o-glycosylated.  
C;Genetics:  
A;Gene: GDB:COL4A2  
A;Map position: 13q34-13q34  
A;Introns: 15/2; 33/3; 1347/1; 1380/1; 1429/1; 1468/1; 1532/1; 1527/3 #status incomplete and subsequently o-glycosylated.  
C;Complex: Type IV collagen is a heterotrimer of two alpha 1(IV) chains (see PIR:CGHU1B) domains (with disulfide and desmosine cross-links), dimeric associations among trimer interrupted helical domain (with disulfide and desmosine cross-links).  
C;Function:  
A;Description: structural component of basement membrane  
C;Superfamily: collagen alpha 2(IX) chain precursor - chicken  
C;Species: Gallus gallus (chicken)  
C;Date: 22-Nov-1993 #sequence\_revision 09-Mar-1996 #text\_change 20-Sep-1999  
C;Accession: S2396; C34493; A28032; B29032; S22062; B28754; A18856; B18856; A28518; R;Niinomiya, Y.; Castagnola, P.; Gerecke, D.; Gordon, M.K.; Jacenko, O.; LuValle, P.; meaguchi, N.; Olsen, B.R.  
A;Title: The molecular biology of collagens with short triple-helical domains. **COL4A2**  
A;Cross-references: EMBL:J02760; NID:9180425; PIDN:AAA58422.1; PID:9180426  
A;Function:  
A;Description: structural component of basement membrane  
C;Superfamily: collagen alpha 2(IX) chain precursor - chicken  
C;Species: Gallus gallus (chicken)  
C;Date: 22-Nov-1993 #sequence\_revision 09-Mar-1996 #text\_change 20-Sep-1999  
C;Accession: S2396; C34493; A28032; MUID:90062114  
A;Status: not compared with conceptual translation  
A;Molecule type: DNA  
A;Residues: 1-677 <NT1>  
R;Niinomiya, I.; Murakami, Y.; Olsen, B.R.  
J. Biol. Chem. 264, 20033-20041, 1989  
A;Title: Tissue-specific forms of type IX collagen-proteoglycan arise from the use of alternative splicing. **COL4A2**  
A;Cross-references: EMBL:J02760; NID:9180425; PIDN:AAA58422.1; PID:9180426  
A;Function:  
A;Description: structural component of basement membrane  
C;Superfamily: collagen alpha 2(IX) chain precursor - chicken  
C;Species: Gallus gallus (chicken)  
C;Date: 22-Nov-1993 #sequence\_revision 09-Mar-1996 #text\_change 20-Sep-1999  
C;Accession: S2396; C34493; A28032; MUID:90062114  
A;Molecule type: mRNA  
A;Residues: 1-174 <NT1>  
A;Cross-references: EMBL:M28660; NID:9211625; PIDN:AAA48709.1; PID:9211626  
R;McCormick, D.; van der Rest, M.; Goodship, J.; Lozano, G.; Niinomiya, Y.; Olsen, B.R.  
Proc. Natl. Acad. Sci. U.S.A. 84, 4044-4048, 1987  
A;Title: Structure of the glycosaminoglycan domain in the type IX collagen-proteoglycan domain. **COL4A2**  
A;Cross-references: EMBL:J02760; NID:9180425; PIDN:AAA58422.1; PID:9180426  
A;Function:  
A;Description: structural component of basement membrane  
C;Superfamily: collagen alpha 2(IX) chain precursor - chicken  
C;Species: Gallus gallus (chicken)  
C;Date: 22-Nov-1993 #sequence\_revision 09-Mar-1996 #text\_change 20-Sep-1999  
C;Accession: S2396; C34493; A28032; MUID:90062114  
A;Molecule type: mRNA  
A;Residues: 120-195 <MC2>  
A;Cross-references: EMBL:J02760; NID:9180425; PIDN:AAA58422.1; PID:9180426  
A;Function:  
A;Description: structural component of basement membrane  
C;Superfamily: collagen alpha 2(IX) chain precursor - chicken  
C;Species: Gallus gallus (chicken)  
C;Date: 22-Nov-1993 #sequence\_revision 09-Mar-1996 #text\_change 20-Sep-1999  
C;Accession: S2396; C34493; A28032; MUID:90062114  
A;Molecule type: protein  
A;Residues: 147-165, 170-180, 'X', 182-188 <MC2>  
R;Lozano, G.; Olsen, B.R.  
A;Cross-references: EMBL:J02760; NID:9180425; PIDN:AAA58422.1; PID:9180426  
A;Function:  
A;Description: structural component of basement membrane  
C;Superfamily: collagen alpha 2(IX) chain precursor - chicken  
C;Species: Gallus gallus (chicken)  
C;Date: 22-Nov-1993 #sequence\_revision 09-Mar-1996 #text\_change 20-Sep-1999  
C;Accession: S2396; C34493; A28032; MUID:90062114  
A;Molecule type: protein  
A;Residues: 401, 'R', 404-631, 'D', 633-677 <LO1>  
A;Cross-references: EMBL:J02760; NID:9180425; PIDN:AAA58422.1; PID:9180426  
R;Lozano, G.; Niinomiya, Y.; Thompson, H.; Olsen, B.R.  
Proc. Natl. Acad. Sci. U.S.A. 82, 4050-4054, 1985  
A;Title: A distinct class of vertebrate collagen genes encodes chicken type IX collagen. **COL4A2**  
A;Cross-references: EMBL:J02760; NID:9180425; PIDN:AAA58422.1; PID:9180426  
A;Function:  
A;Description: structural component of basement membrane  
C;Superfamily: collagen alpha 2(IX) chain precursor - chicken  
C;Species: Gallus gallus (chicken)  
C;Date: 22-Nov-1993 #sequence\_revision 09-Mar-1996 #text\_change 20-Sep-1999  
C;Accession: S2396; C34493; A28032; MUID:90062114  
A;Molecule type: DNA  
A;Residues: 401, 'R', 404-631, 'D', 633-677 <LO1>  
A;Cross-references: EMBL:J02760; NID:9180425; PIDN:AAA58422.1; PID:9180426  
R;Lozano, G.; Niinomiya, Y.; Thompson, H.; Olsen, B.R.  
Proc. Natl. Acad. Sci. U.S.A. 82, 4050-4054, 1985  
A;Title: A distinct class of vertebrate collagen genes encodes chicken type IX collagen. **COL4A2**  
A;Cross-references: EMBL:J02760; NID:9180425; PIDN:AAA58422.1; PID:9180426  
A;Function:  
A;Description: structural component of basement membrane  
C;Superfamily: collagen alpha 2(IX) chain precursor - chicken  
C;Species: Gallus gallus (chicken)  
C;Date: 22-Nov-1993 #sequence\_revision 09-Mar-1996 #text\_change 20-Sep-1999  
C;Accession: S2396; C34493; A28032; MUID:90062114  
A;Molecule type: DNA  
A;Residues: 401, 'R', 404-631, 'D', 633-677 <LO1>  
A;Cross-references: EMBL:J02760; NID:9180425; PIDN:AAA58422.1; PID:9180426  
R;Lozano, G.; Niinomiya, Y.; Thompson, H.; Olsen, B.R.  
Proc. Natl. Acad. Sci. U.S.A. 82, 4050-4054, 1985  
A;Title: A distinct class of vertebrate collagen genes encodes chicken type IX collagen. **COL4A2**  
A;Cross-references: EMBL:J02760; NID:9180425; PIDN:AAA58422.1; PID:9180426  
A;Function:  
A;Description: structural component of basement membrane  
C;Superfamily: collagen alpha 2(IX) chain precursor - chicken  
C;Species: Gallus gallus (chicken)  
C;Date: 22-Nov-1993 #sequence\_revision 09-Mar-1996 #text\_change 20-Sep-1999  
C;Accession: S2396; C34493; A28032; MUID:90062114  
A;Molecule type: DNA  
A;Residues: 401, 'R', 404-631, 'D', 633-677 <LO1>  
A;Cross-references: EMBL:J02760; NID:9180425; PIDN:AAA58422.1; PID:9180426  
R;Lozano, G.; Niinomiya, Y.; Thompson, H.; Olsen, B.R.  
Proc. Natl. Acad. Sci. U.S.A. 82, 4050-4054, 1985  
A;Title: A distinct class of vertebrate collagen genes encodes chicken type IX collagen. **COL4A2**  
A;Cross-references: EMBL:J02760; NID:9180425; PIDN:AAA58422.1; PID:9180426  
A;Function:  
A;Description: structural component of basement membrane  
C;Superfamily: collagen alpha 2(IX) chain precursor - chicken  
C;Species: Gallus gallus (chicken)  
C;Date: 22-Nov-1993 #sequence\_revision 09-Mar-1996 #text\_change 20-Sep-1999  
C;Accession: S2396; C34493; A28032; MUID:90062114  
A;Molecule type: DNA  
A;Residues: 401, 'R', 404-631, 'D', 633-677 <LO1>  
A;Cross-references: EMBL:J02760; NID:9180425; PIDN:AAA58422.1; PID:9180426  
R;Lozano, G.; Niinomiya, Y.; Thompson, H.; Olsen, B.R.  
Proc. Natl. Acad. Sci. U.S.A. 82, 4050-4054, 1985  
A;Title: A distinct class of vertebrate collagen genes encodes chicken type IX collagen. **COL4A2**  
A;Cross-references: EMBL:J02760; NID:9180425; PIDN:AAA58422.1; PID:9180426  
A;Function:  
A;Description: structural component of basement membrane  
C;Superfamily: collagen alpha 2(IX) chain precursor - chicken  
C;Species: Gallus gallus (chicken)  
C;Date: 22-Nov-1993 #sequence\_revision 09-Mar-1996 #text\_change 20-Sep-1999  
C;Accession: S2396; C34493; A28032; MUID:90062114  
A;Molecule type: DNA  
A;Residues: 401, 'R', 404-631, 'D', 633-677 <LO1>  
A;Cross-references: EMBL:J02760; NID:9180425; PIDN:AAA58422.1; PID:9180426  
R;Lozano, G.; Niinomiya, Y.; Thompson, H.; Olsen, B.R.  
Proc. Natl. Acad. Sci. U.S.A. 82, 4050-4054, 1985  
A;Title: A distinct class of vertebrate collagen genes encodes chicken type IX collagen. **COL4A2**  
A;Cross-references: EMBL:J02760; NID:9180425; PIDN:AAA58422.1; PID:9180426  
A;Function:  
A;Description: structural component of basement membrane  
C;Superfamily: collagen alpha 2(IX) chain precursor - chicken  
C;Species: Gallus gallus (chicken)  
C;Date: 22-Nov-1993 #sequence\_revision 09-Mar-1996 #text\_change 20-Sep-1999  
C;Accession: S2396; C34493; A28032; MUID:90062114  
A;Molecule type: DNA  
A;Residues: 401, 'R', 404-631, 'D', 633-677 <LO1>  
A;Cross-references: EMBL:J02760; NID:9180425; PIDN:AAA58422.1; PID:9180426  
R;Lozano, G.; Niinomiya, Y.; Thompson, H.; Olsen, B.R.  
Proc. Natl. Acad. Sci. U.S.A. 82, 4050-4054, 1985  
A;Title: A distinct class of vertebrate collagen genes encodes chicken type IX collagen. **COL4A2**  
A;Cross-references: EMBL:J02760; NID:9180425; PIDN:AAA58422.1; PID:9180426  
A;Function:  
A;Description: structural component of basement membrane  
C;Superfamily: collagen alpha 2(IX) chain precursor - chicken  
C;Species: Gallus gallus (chicken)  
C;Date: 22-Nov-1993 #sequence\_revision 09-Mar-1996 #text\_change 20-Sep-1999  
C;Accession: S2396; C34493; A28032; MUID:90062114  
A;Molecule type: DNA  
A;Residues: 401, 'R', 404-631, 'D', 633-677 <LO1>  
A;Cross-references: EMBL:J02760; NID:9180425; PIDN:AAA58422.1; PID:9180426  
R;Lozano, G.; Niinomiya, Y.; Thompson, H.; Olsen, B.R.  
Proc. Natl. Acad. Sci. U.S.A. 82, 4050-4054, 1985  
A;Title: A distinct class of vertebrate collagen genes encodes chicken type IX collagen. **COL4A2**  
A;Cross-references: EMBL:J02760; NID:9180425; PIDN:AAA58422.1; PID:9180426  
A;Function:  
A;Description: structural component of basement membrane  
C;Superfamily: collagen alpha 2(IX) chain precursor - chicken  
C;Species: Gallus gallus (chicken)  
C;Date: 22-Nov-1993 #sequence\_revision 09-Mar-1996 #text\_change 20-Sep-1999  
C;Accession: S2396; C34493; A28032; MUID:90062114  
A;Molecule type: DNA  
A;Residues: 401, 'R', 404-631, 'D', 633-677 <LO1>  
A;Cross-references: EMBL:J02760; NID:9180425; PIDN:AAA58422.1; PID:9180426  
R;Lozano, G.; Niinomiya, Y.; Thompson, H.; Olsen, B.R.  
Proc. Natl. Acad. Sci. U.S.A. 82, 4050-4054, 1985  
A;Title: A distinct class of vertebrate collagen genes encodes chicken type IX collagen. **COL4A2**  
A;Cross-references: EMBL:J02760; NID:9180425; PIDN:AAA58422.1; PID:9180426  
A;Function:  
A;Description: structural component of basement membrane  
C;Superfamily: collagen alpha 2(IX) chain precursor - chicken  
C;Species: Gallus gallus (chicken)  
C;Date: 22-Nov-1993 #sequence\_revision 09-Mar-1996 #text\_change 20-Sep-1999  
C;Accession: S2396; C34493; A28032; MUID:90062114  
A;Molecule type: DNA  
A;Residues: 401, 'R', 404-631, 'D', 633-677 <LO1>  
A;Cross-references: EMBL:J02760; NID:9180425; PIDN:AAA58422.1; PID:9180426  
R;Lozano, G.; Niinomiya, Y.; Thompson, H.; Olsen, B.R.  
Proc. Natl. Acad. Sci. U.S.A. 82, 4050-4054, 1985  
A;Title: A distinct class of vertebrate collagen genes encodes chicken type IX collagen. **COL4A2**  
A;Cross-references: EMBL:J02760; NID:9180425; PIDN:AAA58422.1; PID:9180426  
A;Function:  
A;Description: structural component of basement membrane  
C;Superfamily: collagen alpha 2(IX) chain precursor - chicken  
C;Species: Gallus gallus (chicken)  
C;Date: 22-Nov-1993 #sequence\_revision 09-Mar-1996 #text\_change 20-Sep-1999  
C;Accession: S2396; C34493; A28032; MUID:90062114  
A;Molecule type: DNA  
A;Residues: 401, 'R', 404-631, 'D', 633-677 <LO1>  
A;Cross-references: EMBL:J02760; NID:9180425; PIDN:AAA58422.1; PID:9180426  
R;Lozano, G.; Niinomiya, Y.; Thompson, H.; Olsen, B.R.  
Proc. Natl. Acad. Sci. U.S.A. 82, 4050-4054, 1985  
A;Title: A distinct class of vertebrate collagen genes encodes chicken type IX collagen. **COL4A2**  
A;Cross-references: EMBL:J02760; NID:9180425; PIDN:AAA58422.1; PID:9180426  
A;Function:  
A;Description: structural component of basement membrane  
C;Superfamily: collagen alpha 2(IX) chain precursor - chicken  
C;Species: Gallus gallus (chicken)  
C;Date: 22-Nov-1993 #sequence\_revision 09-Mar-1996 #text\_change 20-Sep-1999  
C;Accession: S2396; C34493; A28032; MUID:90062114  
A;Molecule type: DNA  
A;Residues: 401, 'R', 404-631, 'D', 633-677 <LO1>  
A;Cross-references: EMBL:J02760; NID:9180425; PIDN:AAA58422.1; PID:9180426  
R;Lozano, G.; Niinomiya, Y.; Thompson, H.; Olsen, B.R.  
Proc. Natl. Acad. Sci. U.S.A. 82, 4050-4054, 1985  
A;Title: A distinct class of vertebrate collagen genes encodes chicken type IX collagen. **COL4A2**  
A;Cross-references: EMBL:J02760; NID:9180425; PIDN:AAA58422.1; PID:9180426  
A;Function:  
A;Description: structural component of basement membrane  
C;Superfamily: collagen alpha 2(IX) chain precursor - chicken  
C;Species: Gallus gallus (chicken)  
C;Date: 22-Nov-1993 #sequence\_revision 09-Mar-1996 #text\_change 20-Sep-1999  
C;Accession: S2396; C34493; A28032; MUID:90062114  
A;Molecule type: DNA  
A;Residues: 401, 'R', 404-631, 'D', 633-677 <LO1>  
A;Cross-references: EMBL:J02760; NID:9180425; PIDN:AAA58422.1; PID:9180426  
R;Lozano, G.; Niinomiya, Y.; Thompson, H.; Olsen, B.R.  
Proc. Natl. Acad. Sci. U.S.A. 82, 4050-4054, 1985  
A;Title: A distinct class of vertebrate collagen genes encodes chicken type IX collagen. **COL4A2**  
A;Cross-references: EMBL:J02760; NID:9180425; PIDN:AAA58422.1; PID:9180426  
A;Function:  
A;Description: structural component of basement membrane  
C;Superfamily: collagen alpha 2(IX) chain precursor - chicken  
C;Species: Gallus gallus (chicken)  
C;Date: 22-Nov-1993 #sequence\_revision 09-Mar-1996 #text\_change 20-Sep-1999  
C;Accession: S2396; C34493; A28032; MUID:90062114  
A;Molecule type: DNA  
A;Residues: 401, 'R', 404-631, 'D', 633-677 <LO1>  
A;Cross-references: EMBL:J02760; NID:9180425; PIDN:AAA58422.1; PID:9180426  
R;Lozano, G.; Niinomiya, Y.; Thompson, H.; Olsen, B.R.  
Proc. Natl. Acad. Sci. U.S.A. 82, 4050-4054, 1985  
A;Title: A distinct class of vertebrate collagen genes encodes chicken type IX collagen. **COL4A2**  
A;Cross-references: EMBL:J02760; NID:9180425; PIDN:AAA58422.1; PID:9180426  
A;Function:  
A;Description: structural component of basement membrane  
C;Superfamily: collagen alpha 2(IX) chain precursor - chicken  
C;Species: Gallus gallus (chicken)  
C;Date: 22-Nov-1993 #sequence\_revision 09-Mar-1996 #text\_change 20-Sep-1999  
C;Accession: S2396; C34493; A28032; MUID:90062114  
A;Molecule type: DNA  
A;Residues: 401, 'R', 404-631, 'D', 633-677 <LO1>  
A;Cross-references: EMBL:J02760; NID:9180425; PIDN:AAA58422.1; PID:9180426  
R;Lozano, G.; Niinomiya, Y.; Thompson, H.; Olsen, B.R.  
Proc. Natl. Acad. Sci. U.S.A. 82, 4050-4054, 1985  
A;Title: A distinct class of vertebrate collagen genes encodes chicken type IX collagen. **COL4A2**  
A;Cross-references: EMBL:J02760; NID:9180425; PIDN:AAA58422.1; PID:9180426  
A;Function:  
A;Description: structural component of basement membrane  
C;Superfamily: collagen alpha 2(IX) chain precursor - chicken  
C;Species: Gallus gallus (chicken)  
C;Date: 22-Nov-1993 #sequence\_revision 09-Mar-1996 #text\_change 20-Sep-1999  
C;Accession: S2396; C34493; A28032; MUID:90062114  
A;Molecule type: DNA  
A;Residues: 401, 'R', 404-631, 'D', 633-677 <LO1>  
A;Cross-references: EMBL:J02760; NID:9180425; PIDN:AAA58422.1; PID:9180426  
R;Lozano, G.; Niinomiya, Y.; Thompson, H.; Olsen, B.R.  
Proc. Natl. Acad. Sci. U.S.A. 82, 4050-4054, 1985  
A;Title: A distinct class of vertebrate collagen genes encodes chicken type IX collagen. **COL4A2**  
A;Cross-references: EMBL:J02760; NID:9180425; PIDN:AAA58422.1; PID:9180426  
A;Function:  
A;Description: structural component of basement membrane  
C;Superfamily: collagen alpha 2(IX) chain precursor - chicken  
C;Species: Gallus gallus (chicken)  
C;Date: 22-Nov-1993 #sequence\_revision 09-Mar-1996 #text\_change 20-Sep-1999  
C;Accession: S2396; C34493; A28032; MUID:90062114  
A;Molecule type: DNA  
A;Residues: 401, 'R', 404-631, 'D', 633-677 <LO1>  
A;Cross-references: EMBL:J02760; NID:918042

A;Accession: A28518  
A;Molecule type: protein  
A;Residues: 116-159, 'X', 161-166, 'X', 168-173, 'X', 175-178 <HUB>  
A;Accession: S23813  
A;Molecule type: protein  
A;Residues: 170-180, 'X', 182-184 <VAN>  
A;Title: The structure of type IX collagen proteoglycan from cartilage is covalently cross-linked to type  
R;Maine, R.; van der Rest, M.; Niomiyama, Y.; Olsen, B.R.  
Ann. N. Y. Acad. Sci. 460, 38-46, 1985  
A;Title: Evidence for aldimine cross-linkage of 190-Lys to collagen alpha 1(II) chain is  
A;Note: evidence for aldimine cross-linkage of 190-Lys to collagen alpha 1(II) chain is  
R;Maine, R.; van der Rest, M.; Niomiyama, Y.; Olsen, B.R.  
A;Molecule type: protein  
A;Residues: 542-567 <MAY>  
C;Genetics:  
A;Intros: 427/3; 439/3; 454/3; 465/3; 514/3; 533/1; 596/1; 622/1  
C;Superfamily: unassigned collagens  
C;Keywords: chondroitin sulfate proteoglycan; coiled coil; extracellular matrix; glycoprotein  
F;1-21/Domain: signal sequence #status predicted <SIG>  
F;22-67/Product: collagen alpha 2(XI) chain #status predicted <MAT>  
F;25-161/Domain: collagenous COL3 #status predicted <COL3>  
F;62-178/Domain: non-collagenous NC3 #status predicted <NC3>  
F;179-517/Domain: collagenous COL2 #status predicted <COL2>  
F;548-547/Domain: non-collagenous NC2 #status predicted <NC2>  
F;648-662/Domain: collagenous COL1 #status predicted <COL1>  
F;663-677/Domain: non-collagenous NC1 #status predicted <NC1>  
F;158-178/Modified site: hydroxyproline (Pro) #status experimental  
F;167-Binding site: chondroitin sulfate (Ser) (covalent) #status experimental  
F;181/Modified site: 5-hydroxylysine (Lys) #status experimental  
F;181/Binding site: carbohydrate (Lys) (covalent) #status experimental  
F;190/Modified site: allysine (Lys) #status predicted  
Query Match Score 63; DB 2; Length 677;  
Best Local Similarity 66.7%; Pred. No. 0.13; 3; Mismatches 10; Conservative 1; Matches 10;  
Qy 1 GVKGKGNPWPAGAP 15  
Db 46 GIDGDKGSPGAPGSP 60

**RESULT 9**

A54849

N;Alternate names: procollagen alpha 1(VII) chain precursor - human  
C;Species: Homo sapiens (man)  
C;Date: 04-Nov-1994 #sequence\_revision 04-Nov-1994 #text\_change 20-Sep-1999  
C;Accession: A54849; PH0844; S16316; 156328; A30296; 184686  
R;Christiano, A.M.; Greenspan, D.S.; Lee, S.; Utton, J.  
J. Biol. Chem. 269, 20266-20262, 1994  
A;Title: Cloning of human type VII collagen. Complete primary sequence of the alpha1(VII) chain  
A;Reference number: A54849; MUID:94327588  
A;Status: not compared with conceptual translation  
A;Molecule type: mRNA  
A;Residues: 1-294 <CHR>  
A;Cross-references: GB:LU2870; NID:9987124; PIDN:AAA7438.1; PID:9987125  
R;Tanaka, T.; Takahashi, K.; Furukawa, F.; Imamura, S.  
Biochem. Biophys. Res. Commun. 183, 958-963, 1992  
A;Title: Molecular cloning and characterization of type VII collagen cDNA.  
A;Reference number: PH0844; MUID:92231902  
A;Accession: PH0844  
A;Molecule type: mRNA  
A;Residues: 'EER', 340-475, 'RALSTASHSTLWRTRHPCNRGSHTWRAEPCPNRASHRAARG', 524-528, 'C'  
A;Cross-references: DDBJ:D11152; DDBJ:D13694; NID:9453698; PIDN:BAA02853.1; PID:9453699  
A;Experimental source: keratinocyte  
A;Note: the authors translated the codon ACC for residues 394 and 397 as Tyr  
R;Parente, M.G.; Chung, L.C.; Ryymänen, J.; Woodley, D.T.; Wynn, K.C.; Bauer, E.A.; Mat  
Proc. Natl. Acad. Sci. U.S.A. 88, 6931-6935, 1991

A;Title: Human type VII collagen: cDNA cloning and chromosomal mapping of the gene.  
A;Reference number: S16316; MUID:9134380  
A;Accession: S16316  
A;Molecule type: mRNA  
A;Residues: 815-892, 'E', 894-1439 <PAR>  
A;Cross-references: GB:MO5158; GB:S49017; NID:9180914; PIDN:AAA96439.1; PID:9180915  
A;Experimental source: keratinocyte  
R;Gammon, W.R.; Abernethy, M.L.; Padilla, K.M.; Prisavanh, P.S.; Cook, M.E.; Wright,  
J. Invest. Dermatol. 99, 691-696, 1992  
A;Title: Noncollagenous (NC1) domain of collagen VII resembles multidomain adhesion protein  
A;Reference number: I56328; MUID:93107742  
A;Accession: I56328  
A;Title: Cleavage of type VII collagen by interstitial collagenase and type IV collagen  
A;Reference number: A30296; MUID:89139437  
A;Accession: A30296  
A;Molecule type: protein  
A;Residues: 'A', 1240-1246, 'G', 1248-1250, 'XE', 1253-1255, 'Q', 1257, 'E', 2032, 'C', 2034-204  
R;Greenspan, D.S.  
Hum. Mol. Genet. 2, 273-278, 1993  
A;Title: The carboxyl-terminal half of type VII collagen, including the non-collagenous  
A;Reference number: I48103; MUID:93271985  
A;Accession: I48103  
A;Status: preliminary; translated from GB/EMBL/DDJB  
A;Molecule type: mRNA  
A;Residues: '2305-2871, 'S', 2873-2944 <RE>  
A;Cross-references: GB:LU6862; NID:938813; PIDN:AAA89196.1; PID:9388714  
R;Christiano, A.M.; Ryymänen, M.; Utton, J.  
Proc. Natl. Acad. Sci. U.S.A. 91, 3549-3553, 1994  
A;Title: Dominant dystrophic epidermolysis bullosa: identification of a GLY --> Ser substitution  
A;Reference number: A55255; MUID:94224777  
A;Comments: annotation  
C;Comments: Prolines and lysines at the third position of the tripeptide repeating unit  
C;Genetics:  
A;Gene: GDB:COL7A1; EBRL; EBD1; EB  
A;Cross-references: GDB:128750; OMIM:120120  
A;Map position: 3p21.3-3p21.3  
A;Note: defects in this gene can result in dominant and recessive dystrophic epidermolysis bullosa  
C;Complex: type VII collagen is probably a homotrimer  
C;Function:  
A;Description: structural component of extracellular polymer associated with anchoring  
C;Superfamily: unassigned collagens; animal Kunitz-type proteinase inhibitor homology  
C;Keywords: coiled coil; extracellular matrix; glycoprotein; hydroxylysine; hydroxyproline  
F;1-16/Domain: signal sequence #status predicted <SIG>  
F;14-16/Product: collagen alpha 1(VII) chain #status predicted <MAT>  
F;17-1253/Domain: amino-terminal nonhelical #status predicted <NC1>  
F;36-201/Domain: von Willebrand factor type A repeat homology <WVAL>  
F;231-318/Domain: fibronectin type III repeat homology <FN1>  
F;327-413/Domain: fibronectin type III repeat homology <FN2>  
F;414-502/Domain: fibronectin type III repeat homology <FN3>  
F;508-593/Domain: fibronectin type III repeat homology <FN4>  
F;598-683/Domain: fibronectin type III repeat homology <FN5>  
F;686-771/Domain: fibronectin type III repeat homology <FN6>  
F;776-862/Domain: fibronectin type III repeat homology <FN7>  
F;864-952/Domain: fibronectin type III repeat homology <FN8>  
F;954-1045/Domain: fibronectin type III repeat homology <FN9>  
F;1052-1219/Domain: von Willebrand factor type A repeat homology <WVA2>  
F;110-112/Region: cell attachment (R-G-D) motif  
F;1189-1233/Region: cysteine/proline-rich  
F;1254-2703/Region: interrupted interrupted  
F;1334-1336/Region: cell attachment (R-G-D) motif  
F;2008-2010/Region: cell attachment (R-G-D) motif  
F;253-2559/Region: cell attachment (R-G-D) motif  
F;2784-2944/Domain: carboxyl-terminal nonhelical #status predicted <NC2>  
F;2876-2929/Domain: animal Kunitz-type proteinase inhibitor homology <BPI>

Qy	1	GVKGDKGNPGWPGAP	15	C; Species: Homo sapiens (man) C; Date: 28-Oct-1994 #sequence_revision 03-Oct-1995 #text_change 22-Jun-1999 C; Accession: A54763; M55928; A44033; A55971; A39786 R; Maruyama, M.; Leinonen, R.; Mochizuki, T.; Tryggvason, K.; Reeder, S.T. J. Biol. Chem. 269, 23013-23017, 1994 A; Title: Complete primary structure of the human alpha3(IV) collagen chain. Coexpressed A; Reference number: A54763; MUID:94364994
Db	2386	GVKGDLGLGGLPGAP	2400	Best Local Similarity 80.0%; Pred. No. 0.55; 3; Indels 0; Gaps 0; Mismatches 0; Status conservative
RESULT 10				Best Local Similarity 80.0%; Pred. No. 0.55; 3; Indels 0; Gaps 0; Mismatches 0; Status conservative
A32249		collagen - sea urchin ( <i>Paracentrotus lividus</i> ) (fragment)		Best Local Similarity 80.0%; Pred. No. 0.55; 3; Indels 0; Gaps 0; Mismatches 0; Status conservative
C;Species:		Paracentrotus lividus (common urchin)		Best Local Similarity 80.0%; Pred. No. 0.55; 3; Indels 0; Gaps 0; Mismatches 0; Status conservative
C;Accession:		A32249		Best Local Similarity 80.0%; Pred. No. 0.55; 3; Indels 0; Gaps 0; Mismatches 0; Status conservative
R;Saitta, B.; Buttice, G.; Gambino, R.				Best Local Similarity 80.0%; Pred. No. 0.55; 3; Indels 0; Gaps 0; Mismatches 0; Status conservative
Bloch, J. Biophys. Res. Commun. 158, 633-639, 1989				Best Local Similarity 80.0%; Pred. No. 0.55; 3; Indels 0; Gaps 0; Mismatches 0; Status conservative
A;Title:		Isolation of a putative collagen-like gene from the sea urchin <i>Paracentrotus lividus</i>		Best Local Similarity 80.0%; Pred. No. 0.55; 3; Indels 0; Gaps 0; Mismatches 0; Status conservative
A;Reference number:		A32249; MUID:89146773		Best Local Similarity 80.0%; Pred. No. 0.55; 3; Indels 0; Gaps 0; Mismatches 0; Status conservative
A;Accession:		A32249		Best Local Similarity 80.0%; Pred. No. 0.55; 3; Indels 0; Gaps 0; Mismatches 0; Status conservative
A;Status:		preliminary; not compared with conceptual translation		Best Local Similarity 80.0%; Pred. No. 0.55; 3; Indels 0; Gaps 0; Mismatches 0; Status conservative
A;Molecule type:		DNA		Best Local Similarity 80.0%; Pred. No. 0.55; 3; Indels 0; Gaps 0; Mismatches 0; Status conservative
C;Keywords:		coiled coil; extracellular matrix; glycoprotein; timer; triple helix		Best Local Similarity 80.0%; Pred. No. 0.55; 3; Indels 0; Gaps 0; Mismatches 0; Status conservative
Query Match		67.4%		Best Local Similarity 80.0%; Pred. No. 0.55; 3; Indels 0; Gaps 0; Mismatches 0; Status conservative
Best Local Similarity		78.6%		Best Local Similarity 80.0%; Pred. No. 0.55; 3; Indels 0; Gaps 0; Mismatches 0; Status conservative
Matches		11; Conservative		Best Local Similarity 80.0%; Pred. No. 0.55; 3; Indels 0; Gaps 0; Mismatches 0; Status conservative
Oy	1	GVKGDKGNPGWPGAP	14	Best Local Similarity 80.0%; Pred. No. 0.55; 3; Indels 0; Gaps 0; Mismatches 0; Status conservative
Db	76	GSKGDQGNPGQPGA	89	Best Local Similarity 80.0%; Pred. No. 0.55; 3; Indels 0; Gaps 0; Mismatches 0; Status conservative
RESULT 11				Best Local Similarity 80.0%; Pred. No. 0.55; 3; Indels 0; Gaps 0; Mismatches 0; Status conservative
S23809		collagen alpha 2(I) chain homolog - sea urchin ( <i>Strongylocentrotus purpuratus</i> )		Best Local Similarity 80.0%; Pred. No. 0.55; 3; Indels 0; Gaps 0; Mismatches 0; Status conservative
C;Species:		Strongylocentrotus purpuratus (purple urchin)		Best Local Similarity 80.0%; Pred. No. 0.55; 3; Indels 0; Gaps 0; Mismatches 0; Status conservative
C;Date:		10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999		Best Local Similarity 80.0%; Pred. No. 0.55; 3; Indels 0; Gaps 0; Mismatches 0; Status conservative
C;Accession:		S23809		Best Local Similarity 80.0%; Pred. No. 0.55; 3; Indels 0; Gaps 0; Mismatches 0; Status conservative
R;Exposito, J.R.; d'Alessio, M.; Solursh, M.; Ramirez, F.				Best Local Similarity 80.0%; Pred. No. 0.55; 3; Indels 0; Gaps 0; Mismatches 0; Status conservative
J. Biol. Chem. 267, 15559-15562, 1992				Best Local Similarity 80.0%; Pred. No. 0.55; 3; Indels 0; Gaps 0; Mismatches 0; Status conservative
A;Title:		Sea urchin collagen evolutionarily homologous to vertebrate pro-alpha-2(I) collagen		Best Local Similarity 80.0%; Pred. No. 0.55; 3; Indels 0; Gaps 0; Mismatches 0; Status conservative
A;Reference number:		S23809; MUID:92340411		Best Local Similarity 80.0%; Pred. No. 0.55; 3; Indels 0; Gaps 0; Mismatches 0; Status conservative
A;Status:		preliminary		Best Local Similarity 80.0%; Pred. No. 0.55; 3; Indels 0; Gaps 0; Mismatches 0; Status conservative
A;Molecule type:		mRNA		Best Local Similarity 80.0%; Pred. No. 0.55; 3; Indels 0; Gaps 0; Mismatches 0; Status conservative
A;Residues:		I-1414 <EXP>		Best Local Similarity 80.0%; Pred. No. 0.55; 3; Indels 0; Gaps 0; Mismatches 0; Status conservative
A;Cross-references:		EMBL:M92040; NID:9161435; PIDN:AAA30035_1; PID:9161436		Best Local Similarity 80.0%; Pred. No. 0.55; 3; Indels 0; Gaps 0; Mismatches 0; Status conservative
C;Superfamily:		Collagen alpha 2(I) chain; fibrillar collagen carboxyl-terminal homology		Best Local Similarity 80.0%; Pred. No. 0.55; 3; Indels 0; Gaps 0; Mismatches 0; Status conservative
C;Keywords:		coiled coil; extracellular matrix; glycoprotein; trimer; triple helix		Best Local Similarity 80.0%; Pred. No. 0.55; 3; Indels 0; Gaps 0; Mismatches 0; Status conservative
Query Match		67.4%		Best Local Similarity 80.0%; Pred. No. 0.55; 3; Indels 0; Gaps 0; Mismatches 0; Status conservative
Best Local Similarity		78.6%		Best Local Similarity 80.0%; Pred. No. 0.55; 3; Indels 0; Gaps 0; Mismatches 0; Status conservative
Matches		11; Conservative		Best Local Similarity 80.0%; Pred. No. 0.55; 3; Indels 0; Gaps 0; Mismatches 0; Status conservative
Oy	1	GVKGDKGNPGWPGAP	14	Best Local Similarity 80.0%; Pred. No. 0.55; 3; Indels 0; Gaps 0; Mismatches 0; Status conservative
Db	393	GSKGDQGNPGQPGA	406	Best Local Similarity 80.0%; Pred. No. 0.55; 3; Indels 0; Gaps 0; Mismatches 0; Status conservative
RESULT 12				Best Local Similarity 80.0%; Pred. No. 0.55; 3; Indels 0; Gaps 0; Mismatches 0; Status conservative
CGH03B		collagen alpha 3(IV) chain precursor, long splice form - human		Best Local Similarity 80.0%; Pred. No. 0.55; 3; Indels 0; Gaps 0; Mismatches 0; Status conservative
N;Alternate names:		Goodpasture antigen; procollagen alpha 3(IV) chain-long splice form		Best Local Similarity 80.0%; Pred. No. 0.55; 3; Indels 0; Gaps 0; Mismatches 0; Status conservative

F;996-998/Region: cell attachment (R-G-D) motif  
 F;154-156/Region: cell attachment (R-G-D) motif  
 F;1306-1308/Region: cell attachment (R-G-D) motif  
 C;Species: Homo sapiens (man)  
 C;Accession: S23810; PQ0612; S08012  
 R;Pan, T.C.; Zhang, R.Z.; Mattei, M.G.; Timpl, R.; Chu, M.L.  
 Proc. Natl. Acad. Sci. U.S.A. 89, 6565-6569, 1992  
 A;Title: Cloning and chromosomal location of human alpha1(XVI) collagen.  
 A;Reference number: S23810; MID:92335339  
 A;Accession: S23810  
 A;Molecule type: mRNA  
 A;Residues: 1-1603 <PAN>  
 A;Cross-references: EMBL:M92642; NID:gi180757; PIDN:AAA58427.1; PID:gi180758  
 A;Experimental source: skin fibroblasts  
 R;Yamaguchi, N.; Kimura, S.; McBride, O.W.; Hori, H.; Yamada, Y.; Kanamori, T.; Yamakoshi, J. Biochem. 112, 856-863, 1992

Query Match Similarity 66.3%; Score 61; DB 1; Length 340;  
 Best Local Similarity 76.9%; Pred. No. 0.12; Matches 10; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 Qy 1 GVKGDKGNPWPWG 13  
 Db 124 GIKGDKGDRGAPG 136

RESULT 14

S23810  
 collagen alpha 1(XVI) chain precursor - human  
 N;Alternate names: procollagen alpha 1(XVI) chain  
 C;Species: Homo sapiens (man)  
 C;Accession: S23810; PQ0612; S08012  
 R;Leinonen, A.; Maruyama, M.; Mochizuki, T.; Tryggvason, K.; Reeder, S.T.  
 J. Biol. Chem. 269, 26172-26177, 1994  
 A;Title: Complete primary structure of the human type IV collagen alpha1(XVI) chain. C  
 A;Reference number: A55360; MID:9501445  
 A;Accession: A55360  
 A;Status: nucleic acid sequence not shown  
 A;Molecule type: mRNA  
 A;Residues: 1-1690 <LEI>  
 A;Cross-references: GB:X81053; NID:g574805; PIDN:CAA56943.1; PID:g574806  
 R;Sugimoto, M.; Oohashi, T.; Yoshioka, H.; Matsuo, N.; Niromiya, Y.

Query Match Similarity 67.4%; Score 62; DB 1; Length 1670;  
 Best Local Similarity 66.7%; Pred. No. 0.44; Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
 Qy 1 GVKGDKGNPWPWG 15  
 Db 1189 GAKGDRGAPGPGLP 1203

RESULT 13

B46345  
 gene V protein - phage PRD1  
 N;Alternate names: minor capsid protein V  
 C;Species: Phage PRD1  
 C;Accession: 17-Feb-1994 #sequence\_revision 17-Feb-1994 #text\_change 23-Jul-1999  
 C;Accession: B46345; D46345  
 R;Bamford, J.K.H.; Bamford, D.H.  
 R;Virology 177, 445-51, 1990  
 A;Title: Capsomeric proteins of bacteriophage PRD1, a bacterial virus with a membrane.  
 A;Reference number: A46345; MID:90320115  
 A;Accession: B46345  
 A;Molecule type: DNA  
 A;Residues: 1-340 <GBAM>  
 A;Cross-references: GB:055568; NID:gi215745; PIDN:AAA32446.1; PID:gi215746  
 A;Accession: B46345  
 A;Molecule type: protein  
 A;Residues: 2-6 <BAZ>  
 C;Genetics:  
 A;Gene: V  
 C;Superfamily: phage PRD1 gene V protein  
 C;Keywords: capsid protein

Query Match Similarity 66.3%; Score 61; DB 1; Length 340;  
 Best Local Similarity 76.9%; Pred. No. 0.12; Matches 10; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 Qy 1 GVKGDKGNPWPWG 13  
 Db 124 GIKGDKGDRGAPG 136

RESULT 14

S23810  
 collagen alpha 1(XVI) chain precursor - human  
 CGHUIB  
 collagen alpha 4 (IV) chain precursor - human  
 N;Alternate names: procollagen alpha 4 (IV) chain  
 C;Species: Homo sapiens (man)  
 C;Accession: S23810; PQ0612; S08012  
 C;Date: 28-Oct-1994 #sequence\_revision 28-Oct-1994 #text\_change 20-Sep-1999  
 C;Accession: S23810; PQ0612; S08012  
 R;Pan, T.C.; Zhang, R.Z.; Mattei, M.G.; Timpl, R.; Chu, M.L.  
 Proc. Natl. Acad. Sci. U.S.A. 89, 6565-6569, 1992  
 A;Title: Cloning and chromosomal location of human alpha1(XVI) collagen.  
 A;Reference number: S23810; MID:92335339  
 A;Accession: S23810  
 A;Molecule type: mRNA  
 A;Residues: 1-1603 <PAN>  
 A;Cross-references: EMBL:M92642; NID:gi180757; PIDN:AAA58427.1; PID:gi180758  
 A;Experimental source: skin fibroblasts  
 R;Yamaguchi, N.; Kimura, S.; McBride, O.W.; Hori, H.; Yamada, Y.; Kanamori, T.; Yamakoshi, J. Biochem. 112, 856-863, 1992

Query Match Similarity 66.3%; Score 61; DB 2; Length 1603;  
 Best Local Similarity 73.3%; Pred. No. 0.59; Matches 11; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 Qy 1 GVKGDKGNPWPWG 15  
 Db 681 GQKGDAGNPFDPGTP 695

RESULT 15

S23810  
 collagen alpha 4 (IV) chain precursor - human  
 CGHUIB  
 collagen alpha 4 (IV) chain precursor - human  
 N;Alternate names: procollagen alpha 4 (IV) chain  
 C;Species: Homo sapiens (man)  
 C;Accession: A55360; S36654; S28777  
 R;Leinonen, A.; Maruyama, M.; Mochizuki, T.; Tryggvason, K.; Reeder, S.T.  
 J. Biol. Chem. 269, 26172-26177, 1994  
 A;Title: Complete primary structure of the human type IV collagen alpha4(IV) chain. C  
 A;Reference number: A55360; MID:9501445  
 A;Accession: A55360  
 A;Status: nucleic acid sequence not shown  
 A;Molecule type: mRNA  
 A;Residues: 1-1690 <LEI>  
 A;Cross-references: GB:X81053; NID:g574805; PIDN:CAA56943.1; PID:g574806  
 R;Sugimoto, M.; Oohashi, T.; Yoshioka, H.; Matsuo, N.; Niromiya, Y.

FBB3 Lett. 330, 122-128, 1993  
A;Title: cDNA isolation and partial gene structure of the human alpha-4(IV) collagen chain  
A;Reference number: S36854; MUID:9374047  
A;Accession: S36854  
A;Molecule type: DNA; mRNA  
A;Cross-references: DDBJ:017391; NID:9440365; PIDN:AAA04214.1; PID:9159649  
A;Experimental source: whole eye  
R;Kamogata, Y.; Mattei, M.G.; Ninomiya, Y.  
J. Biol. Chem. 267, 23753-23758, 1992  
A;Title: Isolation and sequencing of cDNAs and genomic DNAs encoding the alpha4 chain of  
A;Reference number: S28777; MUID:93054733  
A;Accession: S28777  
A;Molecule type: DNA  
A;Residues: 1407-1424, 'G' 1426-1430, 'A' 1432-1439, 'L' 1441-1507 <KAM>  
A;Cross-references: GB:LO1475; GB:LO1476  
A;Note: the codons given for 1438-Asp (GAC) and 1443-Gly (GCA) are inconsistent with the  
C;Comment: Prolines and Lysines at the third position of the tripeptide repeating unit  
ed and subsequently O-glycosylated.  
C;Genetics:  
A;Gene: GDB:COLA4  
A;Cross-references: GDB:132673; OMIM:120131  
A;MM position: 2q35-2q37  
A;Introns: 39/1; 1406/1; 1445/1; 1508/1; 1603/3 #status incomplete  
C;Complex: this minor type IV collagen is thought to form a heterotrimer of two alpha 3(I)  
monomer amino-terminal domains (with disulfide and desmosine cross-links), dimeric  
er associations in the interrupted helical domain (with disulfide and desmosine cross-li  
C;Function:  
A;Description: minor structural component of extracellular basement membrane in kidney q  
C;Superfamily: collagen alpha 1(IV) chain  
C;Keywords: basement membrane; coiled coil; extracellular matrix; glycoprotein; hydroxyl  
F;1-38/Domain: signal sequence #status predicted <SIG>  
F;39-1690/Product: collagen alpha 4(IV) chain #status predicted <MAT>  
F;39-61/Domain: amino-terminal nonhelical, NHI <NHI>  
F;62-466/Region: interrupted helical  
F;94-96/Region: cell attachment (R-G-D) motif  
F;145-147/Region: cell attachment (R-G-D) motif  
F;189-191/Region: cell attachment (R-G-D) motif  
F;310-312/Region: cell attachment (R-G-D) motif  
F;724-726/Region: cell attachment (R-G-D) motif  
F;785-787/Region: cell attachment (R-G-D) motif  
F;989-991/Region: cell attachment (R-G-D) motif  
F;1212-121/Region: cell attachment (R-G-D) motif  
F;1467-1690/Domain: carboxyl-terminal nonhelical, NCI <NCI>  
F;1471-1569/Domain: collagen IV carboxyl-terminal repeat <CTR>  
F;1579-1665/Domain: collagen IV carboxyl-terminal repeat <CTR>  
F;47,52,55,57,266,400,460,492,494,608,790,828,1095,1131,1294,1317,1375,1407/disulfide bc  
F;142,669/binding site: carbohydrate (Asn) (covalent) #status predicted <CTR>  
F;1480-1569,1513-1569/disulfide bonds: (or 1480-1569, 1513-1566) #status predicted  
F;1522-1531,1632-1641/disulfide bonds: #status predicted  
F;1588-1683,1622-1685/disulfide bonds: (or 1588-1686, 1622-1683) #status predicted  
Query Match 66.3%; Score 61; DB 1; Length 1690;  
Best Local Similarity 60.0%; Pred. No. 0.62; Mismatches 4; Indels 0; Gaps 0;  
Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GVKGDKGNPGPAGA 15  
Db 184 GIQGDRODPGLGPGLP 198

RESULT 18  
A45137  
collagen alpha 4(IV) chain precursor - pig roundworm  
C;Species: Ascaris suum (pig roundworm)  
C;Accession: S16366  
C;Accession: 04-Dec-1992 #sequence\_revision 04-Dec-1992 #text\_change 13-Aug-1999  
C;Accession: S16366  
collagen alpha 2(IV) chain precursor - pig roundworm  
C;Species: Ascaris suum (pig roundworm)  
C;Accession: R;Kamogata, Y.; Mattei, M.G.; Ninomiya, Y.  
C;Accession: A45137  
C;Date: 30-Apr-1993 #sequence\_revision 18-Nov-1994 #text\_change 17-Mar-1999  
C;Accession: R;Kamogata, Y.; Mattei, M.G.; Ninomiya, Y.  
J. Biol. Chem. 267, 23753-23758, 1992  
A;Title: Isolation and sequencing of cDNAs and genomic DNAs encoding the alpha4 chain  
A;Reference number: S28777; MUID:93054733  
A;Accession: A45137  
A;Status: preliminary; not compared with conceptual translation  
A;Molecule type: mRNA  
A;Residues: 1-623 <KAM>  
A;Title: The complete primary structure of a nematode alpha-2(IV) collagen and the parti  
A;Reference number: S16366; MUID:9140768  
A;Accession: S16366

A;Molecule type: mRNA  
A;Residues: 1-1763 <KBI>  
A;Cross-references: GB:MG7507; NID:9159648; PIDN:AAA18014.1; PID:9159649  
C;Genetics:  
A;Introns: 229/3; 266/3; 305/3; 360/3; 424/1; 489/1; 548/1; 656/3; 790/1; 891/1; 963/  
C;Superfamily: collagen alpha 1(IV) chain  
C;Keywords: alternative splicing; basement membrane; cell binding; coiled coil; disul  
F;1-26/Domain: signal sequence #status predicted <SIG>  
F;27-173/Domain: non-collagenous alpha 2(IV) chain #status predicted <MAT>  
F;27-172/Domain: non-collagenous NHL #status predicted <NHL>  
F;43-1529/Domain: collagenous #status predicted <COL>  
F;197-199/Region: cell attachment (R-G-D) motif  
F;1530-1763/Domain: carboxyl-terminal nonhelical, NCI #status predicted <NCI>  
F;1639-1763/Domain: repeat NCI #status predicted <NCI1>  
F;31-34,39-41,55-56,53-59/disulfide bonds: interchain #status predicted <CTR>  
F;126/Binding site: carbohydrate (Asn) (covalent) #status predicted  
F;1593-1599,1702-1709/disulfide bonds: #status predicted  
Query Match 66.3%; Score 61; DB 2; Length 1763;  
Best Local Similarity 73.3%; Pred. No. 0.65; Mismatches 3; Indels 0; Gaps 0;  
Matches 11; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GVKGDKGNPGPAGA 15  
Db 760 GLPGMKGNPLPGAP 774

RESULT 17  
A150696  
collagen alpha 1(III) chain - chicken (fragment)  
C;Species: Gallus gallus (chicken)  
C;Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 13-Aug-1999  
C;Accession: I50696  
R;Nah, H.D.; Niu, Z.; Adams, S.L.  
J. Biol. Chem. 269, 16443-16448, 1994  
A;Title: An alternative transcript of the chick type III collagen gene that does not  
A;Reference number: A54041; MUID:94266842  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: mRNA  
A;Residues: 1-310 <NAH>  
A;Cross-references: EMBL:U07974; NID:9520456; PIDN:AAA03409.1; PID:9517433  
A;Genetics:  
A;Gene: CO3J1  
C;Superfamily: collagen alpha 1(1) chain; fibrillar collagen carboxyl-terminal homolo  
Query Match 65.2%; Score 60; DB 2; Length 310;  
Best Local Similarity 71.4%; Pred. No. 0.16; DB 2; Length 310;  
Matches 10; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
Matches 10; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GVKGDKGNPGPAGA 14  
Db 281 GVKGERGSPGPAGA 294

RESULT 18  
A45137  
collagen alpha 4(IV) chain - rabbit  
C;Species: Oryctolagus cuniculus (domestic rabbit)  
C;Accession: R;Kamogata, Y.; Mattei, M.G.; Ninomiya, Y.  
J. Biol. Chem. 267, 23753-23758, 1992  
A;Title: Isolation and sequencing of cDNAs and genomic DNAs encoding the alpha4 chain  
A;Reference number: S28777; MUID:93054733  
A;Accession: A45137  
A;Status: preliminary; not compared with conceptual translation  
A;Molecule type: mRNA  
A;Residues: 1-623 <KAM>  
A;Title: The complete primary structure of a nematode alpha-2(IV) collagen and the parti  
A;Reference number: S16366; MUID:9140768  
A;Accession: S16366

C;Superfamily: collagen alpha 1(IV) chain

Query Match 65.2%; Score 60; DB 2; Length 623;  
 Best Local Similarity 76.9%; Pred. No. 0.32; Matches 10; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Matches 10; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GVKGDKGNPGWPG 13  
 Db 139 GIKDGDKEFSPG 151

---

RESULT 19  
 S41067

collagen alpha 1(III) chain - rat  
 C;Species: Rattus norvegicus (Norway rat)  
 C;Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 13-Aug-1999  
 C;Accession: S41067; A59905; S31924  
 R;Glumoff, V.; Maekela, J.K.; Vuorio, E.  
 Biochim. Biophys. Acta 1217, 41-48, 1994  
 A;Title: Cloning of cDNA for rat pro alpha-1(III) collagen mRNA. Different expression pa  
 A;Reference number: S41067; MUID:94114571  
 A;Accession: S41067  
 A;Status: preliminary  
 A;Molecule type: mRNA  
 A;Residues: 1-636 <PRT>  
 A;Cross-references: EMBL:X70369; NID:957915; PIDN:CAA49832.1; PID:957916  
 R;Frankel, F.R.; Hsu, C.Y.J.; Myers, J.C.; Lin, E.; Lyttle, C.R.; Kohn, B.; Mohn, K.  
 DNA, 7, 347-354, 1988  
 A;Title: Regulation of alpha-2(I), alpha-1 (III), and alpha-2(V) collagen mRNAs by estr  
 A;Reference number: A29905; MUID:88296083  
 A;Accession: A29905  
 A;Molecule type: mRNA  
 A;Residues: 308-482 <PRT>  
 A;Cross-references: GB:A21354; NID:g203500; PIDN:AAA40942.1; PID:g203501  
 R;Gumoff, V.; Maekela, J.K.; Vuorio, E.  
 submitted to the EMBL data library, February 1993  
 A;Reference number: S31924  
 A;Accession: S31924  
 A;Status: preliminary  
 A;Molecule type: mRNA  
 A;Residues: 2-636 <GL2>  
 A;Cross-references: EMBL:X70369  
 C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology:  
 C;Keywords: coiled coil; extracellular matrix; glycoprotein; trimer; triple helix  
 F;465/Domain: fibrillar collagen carboxyl-terminal homology <PCC>

Query Match 65.2%; Score 60; DB 2; Length 636;  
 Best Local Similarity 71.4%; Pred. No. 0.33; Matches 10; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GVKGDKGNPGWPG 14  
 Db 30 GVKGERGSPGGPA 43

---

RESULT 20  
 I5094

collagen alpha 1(III) chain - chicken (fragment)  
 C;Species: Gallus gallus (chicken)  
 C;Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 13-Aug-1999  
 C;Accession: I5094  
 R;Nish, H.D.; Liu, Z.; Adams, S.L. 1994  
 J. Biol. Chem. 269, 16443-16448  
 A;Title: An alternative transcript of the chick type III collagen gene that does not enc  
 A;Reference number: A54041; MUID:94266842  
 A;Accession: I5094  
 A;Status: preliminary; translated from GB/EMBL/DBJ  
 A;Molecule type: mRNA  
 A;Residues: 1-1886 <NAH>  
 A;Cross-references: EMBL:U07973; NID:9520454; PIDN:AAA83407.1; PID:9537432  
 C;Genetics:  
 A;Gene: COL3A1  
 C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology:  
 F;30-90/Domain: von Willebrand factor type C repeat homology <WVC>

---

RESULT 21  
 CGBOVS

collagen alpha 1(III) chain - bovine  
 C;Species: Bos primigenius taurus (cattle)  
 C;Date: 04-Dec-1986 #sequence\_revision 04-Dec-1986 #text\_change 07-May-1999  
 C;Accession: A02867; A38001; A38002; A38003; A38004; A38005; S71946  
 R;Fietzek, P.P.; Altmann, H.; Rauterberg, J.; Heinkel, W.; Wachter, E.; Kuehn, K.  
 Hoppe-Seyler's Z. Physiol. Chem. 360, 809-820, 1979  
 A;Title: The covalent structure of calf skin type III collagen. I. The amino acid seq  
 A;Reference number: A02862; MUID:80026026  
 A;Accession: A02862  
 A;Molecule type: protein  
 A;Residues: 1-242 <PRT>  
 R;Deweis, H.; Fietzek, P.P.; Kuehn, K.  
 Hoppe-Seyler's Z. Physiol. Chem. 360, 821-832, 1979  
 A;Title: The covalent structure of calf skin type III collagen. II. The amino acid se  
 A;Reference number: A38001; MUID:80026027  
 A;Accession: A38001  
 A;Molecule type: protein  
 A;Residues: 243-422 <DENI>  
 R;Benz, H.; Fietzek, P.P.; Kuehn, K.  
 Hoppe-Seyler's Z. Physiol. Chem. 360, 833-840, 1979  
 A;Title: The covalent structure of calf skin type III collagen. III. The amino acid s  
 A;Reference number: A38002; MUID:80026028  
 A;Accession: A38002  
 A;Molecule type: protein  
 A;Residues: 423-571 <BEN>  
 R;Lang, H.; Glanville, R.W.; Fietzek, P.P.; Kuehn, K.  
 Hoppe-Seyler's Z. Physiol. Chem. 360, 841-850, 1979  
 A;Title: The covalent structure of calf skin type III collagen. IV. The amino acid se  
 A;Reference number: A38003; MUID:80026030  
 A;Accession: A38003  
 A;Molecule type: protein  
 A;Residues: 552-808 <LAN>  
 R;Deweis, H.; Fietzek, P.P.; Kuehn, K.  
 Hoppe-Seyler's Z. Physiol. Chem. 360, 851-860, 1979  
 A;Title: The covalent structure of calf skin type III collagen. V. The amino acid seq  
 A;Reference number: A38004; MUID:80026030  
 A;Accession: A38004  
 A;Molecule type: protein  
 A;Residues: 809-947 <DENW>  
 R;Allmann, H.; Fietzek, P.P.; Kuehn, K.  
 Hoppe-Seyler's Z. Physiol. Chem. 360, 861-868, 1979  
 A;Title: The covalent structure of calf skin type III collagen. VI. The amino acid se  
 A;Reference number: A38005; MUID:80026031  
 A;Accession: A38005  
 A;Molecule type: protein  
 A;Residues: 948-1049 <ANL>  
 A;Experimental source: skin  
 R;Heinkel, W.  
 Biochem. J. 318, 497-503, 1996  
 A;Title: Cross-link analysis of the C-telopeptide domain from type III collagen  
 A;Reference number: S71946; MUID:96404897  
 A;Accession: S71946  
 A;Molecule type: protein  
 A;Residues: 857-1061017-1029; 1037-1049 <HEN>  
 C;Comment: Prolines at the third position of the tripeptide repeating unit (G-X-Y) ar  
 C;Comment: The type III collagen molecule is a trimer of identical chains, linked to  
 C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homolo  
 C;Keywords: coiled coil; extracellular matrix; glycoprotein; hydroxylysine; hydroxypro  
 F;1-1049/Product: collagen alpha 1(III) chain #status experimental <CA>  
 F;1-14/Region: amino-terminal nonhelical telopeptide

F;587-589/Region: cell attachment (R-G-D) motif  
F;592-594/Region: cell attachment (R-G-D) motif  
A;Reference number: A90399; MUID:77134724  
A;Accession: A90399  
A;Molecule type: protein  
A;Residues: 'V', 169-225; 229-232, 'P', 234-292, 'D', 294-398 <SEY1>  
F;875-877/Region: cell attachment (R-G-D) motif  
F;877-880/Region: cell attachment (R-G-D) motif  
F;335-937/Region: cell attachment (R-G-D) motif  
F;1041-1049/Region: carboxyl-terminal nonhelical telopeptide  
F;95,107,119,938,950/Modified site: allysine (Lys) #status predicted  
F;107,950/Binding site: carbohydrate (Lys) #status predicted  
F;1040,1041/disulfide bonds: interchain #status predicted

**Query Match** 65.2%; Score 60; DB 1; Length 1049;  
**Best Local Similarity** 71.4%; Pred. No. 0.54; 1; Indels 0; Gaps 0;  
**Matches** 10; **Conservative** 3; **Mismatches** 1; **DB** 702 **Qy** 1 GVKGDKGNPGWPGKA 14  
|||:::||| |||  
**Db** 702 GVKGERSGSPGGPA 715

**RESULT 22**

CGHUTL  
collagen alpha 1(III) chain precursor - human  
A;Alternate names: procollagen alpha 1(III) chain  
C;Species: Homo sapiens (man)  
C;Accession: S05272; S04642; PE0011; S01725; S04887; A90399; A94562; I51868; S59511; A90399  
R;Prockop, D.J.  
submitted to the EMBL Data Library, February 1989  
A;Reference number: S05272  
A;Accession: S05272  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-1240; 'V', 1242-1466 <PRC>  
A;Cross-references: EMBL:X14220; NID:930057; PIDN:CAA32583\_1; PID:930058  
R;Ala-Kokko, L.; Kontusaari, S.; Baldwin, C.T.; Kuivaniemi, H.; Prockop, D.J.  
Biochem. J. 260, 509-516, 1989  
A;Title: Structure of cDNA clones coding for the entire prepro-alpha1(III) chain of human  
A;Reference number: S04642; MUID:89350838  
A;Molecule type: mRNA  
A;Residues: 1-1196 <ALA>  
A;Cross-references: EMBL:X14220; NID:930057; PIDN:CAA32583\_1; PID:930058  
R;Note: the complete sequence is not shown  
R;Henson-Chanda, V.; Su, M.W.; Wall, D.; Chu, M.L.; Ramirez, F.  
Gene 78, 255-265, 1989  
A;Title: Cloning and analysis of the 5' portion of the human type-III procollagen gene  
A;Reference number: PE0011; MUID:89318752  
A;Accession: PE0011  
A;Molecule type: DNA  
A;Residues: 1-176 <EBEN>  
A;Cross-references: GB:M26939; NID:9180813; PIDN:AAA52040\_1; PID:9180814  
R;Toman, P.D.; Rica, G.A.; de Crombrughe, B.  
Nucleic Acids Res. 16, 7201, 1988  
A;Title: Nucleotide sequence of a cDNA coding for the amino-terminal region of human pre  
A;Reference number: S01726; MUID:88303360  
A;Molecule type: mRNA  
A;Residues: 1-170 <TOM>  
A;Cross-references: EMBL:X07240; NID:930060; PIDN:CAA30229\_1; PID:930061  
A;Note: the authors translated the codon CAG for residue 154 as His  
R;Janeczko, R.A.; Ramirez, F.  
Nucleic Acids Res. 17, 6742, 1989  
A;Title: Nucleotide and amino acid sequences of the entire human alpha-1 (III) collagen.  
A;Accession: S04887; MUID:89386015  
A;Accession: S04887  
A;Molecule type: mRNA  
A;Cross-references: EMBL:X07240; NID:930060; PIDN:CAA30229\_1; PID:930061  
R;Janeczko, R.A.; Ramirez, F.  
Nucleic Acids Res. 17, 6742, 1989  
A;Cross-references: EMBL:X15323; NID:929545; PIDN:CAA33387\_1; PID:930045  
A;Note: the authors' translation of residues 905-932 is inconsistent with the nucleotide  
R;Seyer, J.M.; Kang, A.H.  
Biochemistry 16, 158-164, 1977  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Accession: A94562  
A;Residues: 'V', 169-225; 229-277, 'A', 279-292, 'D', 294, 'S', 296-398 <SEY2>  
A;Experimental source: liver  
A;Reference number: A94562  
A;Molecule type: protein  
A;Residues: 'V', 169-225; 229-277, 'A', 279-292, 'D', 294, 'S', 296-398 <SEY1>  
A;Experimental source: liver  
A;Reference number: A94562  
A;Molecule type: protein  
A;Residues: 'V', 169-225; 229-277, 'A', 279-292, 'D', 294, 'S', 296-398 <SEY2>  
A;Experimental source: liver  
A;Reference number: A94562  
A;Note: sequence corrected by A94562; attachment of 2-O-alpha-D-glucosyl-O-beta-D-gal  
R;Seyer, J.M.  
submitted to the Atlas, December 1977  
A;Reference number: A94562  
A;Accession: A94562  
A;Molecule type: protein  
A;Residues: 'V', 169-225; 229-277, 'A', 279-292, 'D', 294, 'S', 296-398 <SEY1>  
A;Experimental source: liver  
A;Note: author submitted corrections to A90399  
R;Milewicz, D.M.; Witz, A.M.; Smith, A.C.; Manchester, D.K.; Waldstein, G.; Byers, P.  
Am. J. Hum. Genet. 53, 62-70, 1993  
A;Title: Parental somatic and germ-line mosaicism for a multiexon deletion with unusu  
A;Reference number: 151868; MUID:93304430  
A;Accession: 151868  
A;Status: preliminary; translated from GB/EMBL/DDJB  
A;Molecule type: DNA  
A;Residues: 186-194 <ML>  
A;Cross-references: GB:S62925; NID:9386425; PIDN:AAD13937\_1; PID:94261637  
R;Chiodo, A.A.; Silience, D.O.; Cole, W.G.; Bateman, J.F.  
Biochem. J. 311, 939-943, 1995  
A;Title: Abnormal type III collagen produced by an exon-17-skipping mutation of the C  
A;Reference number: S59511; MUID:96067614  
A;Accession: S59511  
A;Molecule type: mRNA  
A;Residues: 302-423 <CH1>  
A;Cross-references: GB:S79877; NID:9195576; PIDN:AAB35615\_1; PID:9195576  
A;Molecule type: protein  
A;Residues: 399-675; 'N', 677-727 <SEY3>  
A;Experimental source: liver  
R;Lee, B.; Vitale, B.; Superti-Furga, A.; Steinmann, B.; Ramirez, F.  
J. Biol. Chem. 266, 5226-5229, 1991  
A;Title: Covalent structure of collagen: amino acid sequence of five consecutive CNBr  
A;Reference number: A90414; MUID:7900043  
A;Accession: A90414  
A;Molecule type: protein  
A;Residues: 399-675; 'N', 677-727 <SEY3>  
A;Experimental source: liver  
R;Lee, B.; Vitale, B.; Superti-Furga, A.; Steinmann, B.; Ramirez, F.  
J. Biol. Chem. 266, 5226-5229, 1991  
A;Title: G to T transversion at position +5 of a splice donor site causes skipping of  
A;Reference number: 155349; MUID:91161621  
A;Accession: 155349  
A;Status: translated from GB/EMBL/DDJB  
A;Molecule type: DNA  
A;Residues: 537-605 <LER>  
A;Cross-references: GB:M59312; NID:9180815; PIDN:AAA52041\_1; PID:9180816  
R;Seyer, J.M.; Mainardi, C.; Kang, A.H.  
Biochemistry 19, 1583-1589, 1980  
A;Title: Covalent structure of collagen: amino acid sequence of alpha1 (III)-CB5 from  
A;Reference number: A90438; MUID:80198282  
A;Accession: A90438  
A;Molecule type: protein  
A;Residues: 728-895; 'A', 897-964 <SEY4>  
A;Experimental source: liver  
R;Cole, W.G.; Chiodo, A.A.; Lamande, S.R.; Janeczko, R.; Ramirez, F.; Dahl, H.H.M.; C  
J. Biol. Chem. 265, 17070-17077, 1990  
A;Title: A base substitution at a splice site in the COL3A1 gene causes exon skipping  
A;Reference number: A38303; MUID:91009133  
A;Accession: A38303  
A;Molecule type: mRNA  
A;Residues: 801-1015 <COL>  
A;Cross-references: GB:J05617; GB:N55603; GB:M59227; NID:9180878; PIDN:ABA59383\_1; P  
R;Mankoo, B.S.; Daigleish, R.  
Nucleic Acids Res. 16, 2337, 1988  
A;Title: Human pro alpha1(III) collagen: cDNA sequence for the 3' end.  
A;Reference number: S02119; MUID:88189827  
A;Accession: S02119  
A;Status: translation not shown

A;Molecule type: mRNA  
A;Residues: 950-1018,'Y',1020-1183,'S',1185-1466 <MAN>  
A;Cross-references: EMBL:X06700; NID:g30053; PIDN:CAA29886\_1; PID:g30054  
R;Sever, J.M.; Kang, A.H.  
Biochemistry 20, 2621-2627, 1981  
A;Title: Covalent structure of collagen: amino acid sequence of alpha1 (III)-CB9 from ty  
A;Reference number: A90446; MUID:81208139  
A;Accession: A90446  
A;Molecule type: protein  
A;Residues: 965-979,'A',981-984,'PS',987,'QN',990-1095,'P',1098-1152,'AT',1155,'S',1157-  
A;Experimental source: Liver  
R;Loidl, H.R.; Brinker, J.M.; May, M.; Pihlajaniemi, T.; Morrow, S.; Rosenblum, J.; Mye  
Nucleic acids Res. 12, 9383-9394, 1984  
A;Title: Molecular cloning and carboxyl-terminal propeptide analysis of human type III procollage  
A;Reference number: A93551; MUID:85087944  
A;Molecule type: mRNA  
A;Residues: 1155,'P',1157-1466 <LOI>  
A;Cross-references: EMBL:X01655; EMBL:X01742; NID:g29884; PIDN:CAA25021\_1  
R;Miskulin, M.; Daigle, R.; Kljuve-Beckerhan, B.; Rennard, S.I.; Tolstoshev, P.; Brant  
Biochemistry 25, 1408-1413, 1986  
A;Title: Human type III collagen gene expression is coordinately modulated with the type  
A;Reference number: I52393; MUID:86187804  
A;Accession: I52393  
A;Status: translated from GB/EMBL/DDBJ  
A;Molecule type: mRNA  
A;Residues: 1161-1200 <MTS>  
A;Cross-references: GB:MI3146; NID:g180415; PIDN:AAA52003\_1; PID:g180416  
R;Emanuel, B.S.; Cannizaro, L.A.; Sevier, J.M.; Myers, J.C.  
Proc. Natl. Acad. Sci. U.S.A. 82, 3385-3389, 1985  
A;Title: Human alpha 1(III) and alpha 2(V) procollagen genes are located on the long arm  
A;Reference number: I59025; MUID:85216505  
A;Accession: I779359  
A;Status: translated from GB/EMBL/DDBJ  
A;Molecule type: mRNA  
A;Residues: 1165-1196 <EMA>  
A;Cross-references: GB:MI1134; NID:g180417; PIDN:AAA50004\_1; PID:g180418  
R;Chu, M.L.; Weil, D.; de Wet, W.; Bernard, M.; Sippola, M.; Ramirez, F.  
J. Biol. Chem. 260, 4357-4363, 1985  
A;Title: Isolation of cDNA and genomic clones encoding human pro-alpha1(III) collagen. F  
A;Reference number: A92516; MUID:85157600  
A;Accession: A92516  
A;Molecule type: DNA  
A;Residues: 1176-1240,'V',1242-1356,'P',1358-1466 <CHI>  
A;Cross-references: GB:MI0615; GB:MI0793; GB:MI0794; GB:MI0795; GB:MI0796; GB:MI0797; GE  
A;Experimental source: Liver  
A;Title: The authors translated the codon TTC for residue 1057 as Tyr; the codons given f  
A;Comment: Prolines and lysines at the third position of the tripeptide repeating unit C  
C;Comment: About 15% of the lysines are 5-hydroxylated and some are subsequently C  
C;Genetics:  
A;Gene: GJB:COL3A1  
A;Cross-references: GDB:118729; OMIM:120180  
A;Map position: 2q31-2q31.1  
A;Note: the list of introns is incomplete; defects in this gene can result in Ehlers-Dan  
C;Complex: type III collagen is a homotrimer of monomers initially linked by disulfide b  
er of their length, is formed with desmosine cross-links made from lysine and allysine r  
C;Function:  
A;Description: structural component of extracellular fibrous polymer that maintains inte  
C;Superfamily: collagen alpha1(II) chain; fibrillar collagen carboxyl-terminal homology; hyd  
C;Keywords: coiled coil; Ehlers-Danlos syndrome; extracellular matrix; glycoprotein; hyd  
F;1-23/Domain: signal sequence #status predicted <SIG>  
F;2-4,153/Domain: amino-terminal propeptide #status predicted <PRO>  
F;31-91/Domain: von Willebrand factor type C repeat homology <VWC>  
F;154-1221/Product: collagen alpha1(III) chain #status predicted <MAT>  
F;155-167/Region: amino-terminal nonhelical telopeptide  
F;168-1196/Region: helical  
F;1091-1093/Region: cell attachment (R-G-D) motif  
F;1197-1221/Region: carboxyl-terminal nonhelical telopeptide  
F;1222-1466/Domain: carboxyl-terminal propeptide #status predicted <CP>  
F;1238-1466/Domain: fibrillar collagen carboxyl-terminal homology <FC>  
F;24/Modified site: pyrrolidone carboxylic acid (Gin) (in mature form) #status predicted

F;153-154/Cleavage site: Pro-Gln (procollagen N-endopeptidase) #status predicted  
F;154/Modified site: pyrrolidone carboxylic acid (Gin) (in mature form) #status pred  
F;161-1212/Modified site: allysine (Lys) #status predicted  
F;63,-2,24,850,-97,-1106/Modified site: 5-hydroxylysine (Lys) #status experimental  
F;263/Binding site: carbohydrate (Lys) (covalent) #status experimental  
F;948-949/Cleavage site: Gly-ile (collagenase) #status experimental  
F;1106/Binding site: carbohydrate (Lys) (covalent) #status predicted  
F;584-1094/Modified site: 5-hydroxylysine (Lys) (partial) #status experimental  
Query Match 65 2%; Score 60; DB 1; Length 1466;  
Best Local Similarity 78.6%; Pred. No. 0.76; DB 1; Length 1466;  
Matches 11; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 1 GVKGDKGNPGWPGA 14  
Db 738 GPKGDKGEPGCPGPA 751  
RESULT 23  
A61262  
A;Cross-references: GB:MI1669; NID:g179516; PIDN:AAA35605\_1; PID:g179517  
R;Li, K.H.; Sawamura, D.; Giudice, G.J.; Diaz, L.A.; Mattei, M.G.; Chu, M.L.; Ulti  
J. Biol. Chem. 266, 2406-2409, 1991  
A;Title: Genomic organization of collagenous domains and chromosomal assignment of hu  
A;Reference number: I55345; MUID:92084712  
A;Accession: I55345  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: mRNA  
A;Residues: 1-1532 <RES>  
A;Cross-references: GB:MI1669; NID:g179516; PIDN:AAA35605\_1; PID:g179517  
R;Giudice, G.J.; Emery, D.; Diaz, L.A.  
J. Invest. Dermatol. 99, 243-250, 1992  
A;Title: Cloning and primary structural analysis of the Bullous pemphigoid autoantige  
A;Reference number: I56325; MUID:92381323  
A;Accession: I56325  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: mRNA  
A;Residues: 1-1532 <RES>  
A;Cross-references: GB:MI1669; NID:g179516; PIDN:AAA35605\_1; PID:g179517  
R;Giudice, G.J.; Squierra, H.L.; Elias, P.M.; Diaz, L.A.  
J. Clin. Invest. 87, 734-738, 1991  
A;Title: Identification of two collagen domains within the bullous pemphigoid autoant  
A;Reference number: A61262; MUID:9123476  
A;Accession: A61262  
A;Molecule type: mRNA  
A;Residues: 543-890,'P', <RE2>  
C;Genetics:  
A;Gene: GJB:COL1A1; BPAG2; BP180  
A;Cross-references: GDB:131396; OMIM:113811  
A;Map position: 10q24.3-11q24.3  
Query Match 65.2%; Score 60; DB 2; Length 1532;  
Best Local Similarity 78.6%; Pred. No. 0.79; DB 2; Length 1532;  
Matches 11; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
QY 1 GVKGDKGNPGWPGA 14  
Db 995 GPKGDKGEPGCPGPA 1008  
RESULT 24  
T13990  
Collagen type IV alpha 2 - fruit fly (*Drosophila melanogaster*)  
C;Species: *Drosophila melanogaster*  
C;Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 17-Nov-2000  
C;Accession: T13990  
R;Yasothornskri, S.; Davis, W.J.; Cramer, G.; Kimbrell, D.A.; Dearolf, C.R.  
submitted to the EMBL Data Library, July 1996  
A;Description: Viking: identification and characterization of a novel type IV collage

A: Reference number: 217845  
A: Accession: TI390  
A: Status: preliminary; translated from GB/EMBL/DDJB  
A: Molecule type: mRNA  
A: Residues: 1-1761 <XAS>  
A: Cross-references: EMBL:U65431; NID:92281290; PID:92281291; PIDN:AAB64082.1  
C: Genetics:  
A: Gene: Cola2  
A: Cross-references: FlyBase:FBgn0016075  
C: Superfamily: collagen alpha 1(IV) chain

Query Match 65.2%; Score 60; DB 2; Length 1761;  
Best Local Similarity 66.7%; Pred. No. 0.91; 2; Mismatches 3; Indels 0; Gaps 0;  
Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GVGDKGNPGWGCAP 15  
Db 994 GRIGETENPGFGRP 1008

RESULT 25

A45407  
Collagen alpha 3(IV) chain - sea urchin (*Strongylocentrotus purpuratus*)  
C: Date: 22-Sep-1993 #sequence\_revision 18-Nov-1994 #text\_change 13-Aug-1999  
C: Accession: A45407; A43903; A23940  
R: Exposito, J.Y.; D'Alessandro, M.; Di Liberto, M.; Ramirez, F.  
J. Biol. Chem. 266, 5349-5354, 1993  
A: Title: Complete primary structure of a sea urchin type IV collagen alpha chain and analysis  
A: Reference number: A45407; MUID:93186842  
A: Accession: A45407  
A: Status: preliminary; not compared with conceptual translation  
A: Molecule type: nucleic acid  
A: Residues: 1-1752 <EXP>  
A: Note: sequence extracted from NCBI backbone (NCBIP:126841)  
R: Wessel, G.M.; Etkin, M.; Benson, S.  
Dev. Biol. 148, 261-272, 1991  
A: Title: Primary mesenchyme cells of the sea urchin embryo require an autonomously produced  
A: Reference number: A43903; MUID:92038439  
A: Accession: A43903  
A: Status: preliminary  
A: Molecule type: mRNA  
A: Residues: 'P', 633->537, 'G', <WNS>  
A: Cross-references: GB:S64572; PIDN:AAE20270.1; PID:9238617  
A: Note: sequence extracted from NCBI backbone (NCBIM:64572; NCBIP:64573)  
A: Molecule type: DNA  
A: Residues: 742-812 <EN>  
A: Cross-references: EMBL:M13206  
A: Molecule type: DNA  
C: Keywords: collagen alpha 1(IV) chain  
F162-161-/Domain: amino-terminal nonhelical, 7S <7SD>  
F162-1523/Region: interrupted helical  
F1524-1752/Domain: carboxyl-terminal nonhelical, NC1 <NC1>  
F1534-1634/Domain: collagen IV carboxyl-terminal repeat <CT1>  
F1644-1748/Domain: collagen IV carboxyl-terminal repeat <CT2>  
F1719/Modified site: allysine (Lys) #status predicted  
Query Match 64.7%; Score 59.5; DB 2; Length 1752;  
Best Local Similarity 54.2%; Pred. No. 1.1; 2; Mismatches 13; Conservative 0; Mismatches 2; Indels 9; Gaps 1;

Qy 1 GVGDKGNPGWGCAP 15  
Db 1177 GVGDPGRTGPEGAKGNGLPGIP 1200

RESULT 26

S33603

surfactant protein D - bovine  
C: Species: Bos primigenius taurus (cattle)  
C: Date: 02-Dec-1993 #sequence\_revision 01-Sep-1995 #text\_change 17-Mar-1999  
C: Accession: S33603  
R: Lim, B.L.; Lu, J.; Reid, K.B.M.  
Immunology 78, 159-165, 1993  
A: Title: Structural similarity between bovine conglutinin and bovine lung surfactant  
A: Reference number: S33603; MUID:93170856  
A: Status: preliminary  
A: Molecule type: mRNA  
A: Residues: 1-369 <CLIM>  
C: Superfamily: pulmonary surfactant protein D; C-type lectin homology <LCH>  
F:248-367/Domain: C-type lectin homology <LCH>  
Query Match 64.1%; Score 59; DB 2; Length 369;  
Best Local Similarity 66.7%; Pred. No. 0.27; 2; Mismatches 3; Indels 0; Gaps 0;  
Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GVGDKGNPGWGCAP 15  
Db 157 GLKGERGAPGDPGAP 171

RESULT 27

A39024  
Collagen alpha 3(IV) chain - bovine (fragment)  
C: Species: Bos primigenius tauris (cattle)  
C: Date: 04-Dec-1992 #sequence\_revision 04-Dec-1992 #text\_change 13-Aug-1999  
C: Accession: A39024; S20672; S17802; A3567; C39419; S1747; S20815  
R: Morrison, K.E.; Germino, G.G.; Reeders, S.T.  
J. Biol. Chem. 261, 7074-7077, 1987  
A: Title: Use of the polymerase chain reaction to clone and sequence a cDNA encoding the Goodpasture epitope to a novel chain of basement membrane  
A: Reference number: A39024; MUID:9103146  
A: Accession: A39024  
A: Molecule type: mRNA  
A: Residues: 1-471 <MOR>  
A: Cross-references: EMBL:MG3139; NID:9162886; PIDN:AAA62708.1; PID:9162887  
A: Note: sequence extracted from NCBI backbone (NCBIM:64572; NCBIP:64573)  
R: Butkowski, R.J.; Langeveld, J.P.M.; Wieslander, J.; Hamilton, J.; Hudson, B.G.  
J. Biol. Chem. 262, 7074-7077, 1987  
A: Title: Localization of the Goodpasture antigen as the alpha-3(IV) chain of basement membrane  
A: Reference number: S18432; MUID:87222419  
A: Accession: S18432  
A: Molecule type: protein  
A: Residues: 227-228, 'X', 230-244 <BUT>  
R: Saus, J.; Wieslander, J.; Langeveld, J.P.M.; Quinones, S.; Hudson, B.G.  
J. Biol. Chem. 263, 1374-1380, 1988  
A: Title: Identification of the Goodpasture antigen as the alpha-3(IV) chain of collagen  
A: Reference number: S17802; MUID:88330844  
A: Accession: S17802  
A: Molecule type: protein  
A: Residues: 227-228, 'X', 230-252, 'Y', 254 <SAU>  
R: Gunwar, S.; Saus, J.; Noelken, M.E.; Hudson, B.G.  
J. Biol. Chem. 265, 5466-5469, 1990  
A: Title: Glomerular basement membrane. Identification of a fourth chain, alpha4, of the basement membrane. Identification of dimeric subunits of the noncollagenous protein  
A: Reference number: A35167; MUID:90202779  
A: Accession: A35167  
A: Molecule type: protein  
A: Residues: 236-258 <GUN>  
R: Gunwar, S.; Ballester, F.; Kalluri, R.; Timoneda, J.; Chonko, A.M.; Edwards, S.J.; J. Biol. Chem. 266, 15138-15324, 1991  
A: Title: Glomerular basement membrane. Identification of dimeric subunits of the noncollagenous protein  
A: Reference number: A39419; MUID:91332055  
A: Accession: C39419  
A: Molecule type: protein  
A: Residues: 236-255 <GUN>  
C: Superfamily: collagen alpha 1(IV) chain  
C: Keywords: basement membrane; cell binding; coiled coil; disulfide bond; duplication  
F1-238/Domain: collagen basement membrane. Identification of dimeric subunits of the noncollagenous protein  
F1-239-471/Domain: carboxyl-terminal nonhelical, NC1 #status predicted <NC1>  
F1-239-473/Domain: repeat NC1 #status predicted <NC1>  
F:232,238/Modified site: hydroxyproline (Pro) #status experimental

F;306-312,417-423/Disulfide bonds: #status predicted	Query Match 64.1%; Score 59; DB 2; Length 471;
Best Local Similarity 66.7%; Pred. No. 0.34; Mismatches 3; Indels 0; Gaps 0;	A: Reference number: A34476; MUID:90008929
Matches 10; Conservative 2; Misnatches 3; Indels 0; Gaps 0;	A: Accession: B34476
Qy 1 GVKGDKNPQWGP 15	A: Molecule type: DNA
RESULT 28	A: Residues: 1432-1499; 'Q' 1501-1707; 'P' 1709-1744 <GU2>
Db 116 GNKGKGNNSGFPGPP 130	C:Keywords: basement membrane; cell binding; coiled coil; disulfide bond; duplication
	F:43-1515/Domain: collagenous, triple helix # status predicted <COL>
	F:93-95/Region: cell attachment (R-G-D) motif
	F:1053-1055/Region: cell attachment (R-G-D) motif
	C:Species: Mus musculus (house mouse)
	C:Date: 21-Sep-1993 #sequence_revision 21-Sep-1993 #text_change 05-Nov-1999
	A:Accession: A46053
	R:Lii, K.; Tamai, K.; Tan, E.M.L.; Uitto, J.
	J. Biol. Chem. 268, 8825-8834, 1993
	A:Title: Cloning of type XVII collagen. Complementary and genomic DNA sequences of mouse tailous pemphigoid antigen, BPAG2, type XVII collagen alpha 1-chain - mouse
	A:Reference number: A46053; MUID:93232041
	A:Status: preliminary
	A:Molecule type: mRNA
	A:Residues: 1-1433 <LL1>
	A:Cross-references: GB:LP08407; PIDN:AAA37443.1; PID:9309183
	A:Note: sequence extracted from NCBI backbone (NCBIN:129627, NCBIPI:129628)
Query Match 64.1%; Score 59; DB 2; Length 1433;	Best Local Similarity 66.7%; Pred. No. 1; Mismatches 2; Indels 0; Gaps 0;
Matches 10; Conservative 2; Misnatches 3; Indels 0; Gaps 0;	Qy 1 GVKGDKNPQWGP 15
Db 948 GPKGDDDPGVPGTP 962	A:Accession: S40991
RESULT 29	A:Description: The sequence of C. elegans cosmid K03H9.
S40991	A:Reference number: S60597
collagen alpha 1(IV) chain precursor - Caenorhabditis elegans	A:Accession: S60598
C:Species: Caenorhabditis elegans	A:Cross-references: EMBL:U21318; PIDN:9687828; PIDN:AC46668.1; PID:9687830
C:Accession: S40991; S4442; S13651; B34476	A:Residues: 1-1285 <AND>
R:Ainscough, R.; Ainscough, R.; Kramer, J.M.	A:Accession: T16549
submitted to the EMBL Data Library, October 1993	A:Cross-references: EMBL:U21318; PIDN:9687828; PID:9687830; PIDN:AC46668.1; CESP:K03H
A:Accession: S40991	A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA	A:Residues: 1-285 <AN2>
A:Residues: 1-174 <AIN>	A:Accession: T16549
A:Cross-references: EMBL:227078; PID:9414627; PID:9414628	A:Cross-references: strain Bristol N2
R:Kramer, J.M.	A:Experimental source: rasper, A.; Fulton, L.; Gardner, A.; Green, P.; Hawkins, T.; Hillier, L.; Jain, M.; Burto
submitted to the EMBL Data Library, December 1990	R:Wilson, R.; Ainscough, R.; Anderson, K.; Baynes, C.; Berks, M.; Bonfield, J.; Burto
A:Reference number: S4442	R:Wilson, R.; Ainscough, R.; Anderson, K.; Baynes, C.; Berks, M.; Bonfield, J.; Burto
A:Accession: S4442	R:Wilson, R.; Ainscough, R.; Anderson, K.; Baynes, C.; Berks, M.; Bonfield, J.; Burto
A:Molecule type: DNA	R:Wilson, R.; Ainscough, R.; Anderson, K.; Baynes, C.; Berks, M.; Bonfield, J.; Burto
A:Residues: 1-129 'GFPGMPGLAGPQQGQNGNPRGPLSQQPGEQGVNSQGRKGVKGESGRSGVPGLP'	R:Wilson, R.; Ainscough, R.; Anderson, K.; Baynes, C.; Berks, M.; Bonfield, J.; Burto
Nature 349, 707-709, 1991	R:Wilson, R.; Ainscough, R.; Anderson, K.; Baynes, C.; Berks, M.; Bonfield, J.; Burto
A:Reference number: S13651; MUID:9141582	R:Wilson, R.; Ainscough, R.; Anderson, K.; Baynes, C.; Berks, M.; Bonfield, J.; Burto
A:Cross-references: EMBL:X56979; PID:96675; PIDN:CA440299.1; PID:96676	A:Authors: Showenkeen, R.; Sims, M.; Smaldo, N.; Smith, A.; Smith, M.; Sonnhammer, E.
R:Guo, X.; Johnson, J.J.; Kramer, J.M.	A:Title: 2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans.
A:Accession: S4442	A:Reference number: S33531; MUID:94150718
A:Molecule type: DNA	A:Contents: annotation
A:Residues: 1-129 'GFPGMPGLAGPQQGQNGNPRGPLSQQPGEQGVNSQGRKGVKGESGRSGVPGLP'	C:Genetics:
'N, 'D, 817-1260, 'P, 1262-1707, 'P, 1709-1744 <KRA>	A:Gene: CESP:K03H.2
A:Accession: S13651	A:Map position: II
A:Cross-references: EMBL:X56979	A:Introns: 93/1; 147/3; 247/1
A:Status: nucleic acid sequence not shown	C:Superfamily: unassigned collagens
A:Molecule type: DNA	Query Match 63.0%; Score 58; DB 2; Length 285;
	Best Local Similarity 60.0%; Pred. No. 0.29; Mismatches 5; Indels 0; Gaps 0;
	Qy 1 GVKGDKNPQWGP 15
	Db 142 GIDGPGIPGWGP 156

Fri Nov 1 13:09:20 2002

us-09-529-691a-1.rpr

Page 17

Job time : 15.5 secs

**THIS PAGE BLANK (uspto)**

Gencore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using SW model  
Run on: November 1, 2002, 12:16:06 ; Search time 7 Seconds  
Perfect score: US-09-529-691a-1  
Sequence: 1 GVKGKGKGNPGWPGAP 15

Title: US-09-529-691a-1  
Post-processing: Maximum Match 0%  
Database : Swissprot\_40;\*

Scoring table: BLOSUM62  
Gapop 11.0 , Gapext 0.5

Searched: 105224 seqs, 3879550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Maximum Match 0%

Listing first 45 summaries

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

**SUMMARIES**  
\*  

Result No.	Score	Query Match Length	DB ID	Description
1	92	100.0	1669	1 CA14_HUMAN
2	78	84.8	1669	1 CA14_MOUSE
3	69	75.0	1685	1 CA5_HUMAN
4	66	71.7	1754	1 CA54_CANFA
5	65	70.7	1707	1 CA24_MOUSE
6	64	69.6	1712	1 CA2_HUMAN
7	63	68.5	2944	1 CA17_HUMAN
8	62	67.4	1670	1 CA4_HUMAN
9	61	66.3	339	1 COA5_BPRD
10	61	66.3	1603	1 CA1F_HUMAN
11	61	66.3	1690	1 CA44_HUMAN
12	61	66.3	1763	1 CA24_ASCSU
13	60	65.2	369	1 PSPD_BOVIN
14	60	65.2	623	1 CA4_RABIT
15	60	65.2	636	1 CA13_RAT
16	60	65.2	1049	1 CA11_BOVIN
17	60	65.2	1262	1 CA13_CHICK
18	60	65.2	1465	1 CA13_HUMAN
19	59	64.1	471	1 CA34_BOVIN
20	59	64.1	1758	1 CA14_CAEEL
21	58	63.0	285	1 YRF2_CAEEL
22	58	63.0	1516	1 CA1H_HUMAN
23	58	63.0	1758	1 CA24_CAEEL
24	57	62.0	298	1 CC34_CAEEL
25	57	62.0	483	1 MRCO_MESAU
26	57	62.0	1356	1 CA21_ONCMY
27	56	60.9	518	1 MTCO_MOUSE
28	56	60.9	1143	1 CA1L_HUMAN
29	55	60.9	1464	1 CA13_MOUSE
30	55	60.3	458	1 MSRE_MOUSE
31	55	59.8	245	1 C1OC_HUMAN
32	55	59.8	266	1 YWK_CAEEL
33	55	59.8	371	1 CONG_BOVIN

**ALIGNMENTS**  

RESULT 1	CA14_HUMAN	STANDARD:	PRT:	1669 AA.
RP	ID: CA14_HUMAN			
RC	SEQUENCE OF 46-1257 FROM N.A.			
RX	TISSUE=Placenta;			
RX	MEDLINE=88083584; PubMed=3691802;			
RA	Soininen R., Haka R., Risku T., Prockop D.J., Tryggvason K.;			
RT	"Complete primary structure of the alpha 1-chain of human basement membrane (type IV) collagen";			
DT	01-FEB-1996 (Rel. 01, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Collagen alpha 1(IV) chain precursor.			
GN	COL4A1.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=89340433; PubMed=2701944;			
RA	Soininen R., Huotari M., Ganguly A., Prockop D.J., Tryggvason K.;			
RT	"Structural organization of the gene for the alpha 1 chain of human type IV collagen";			
RT	J. Biol. Chem. 264:13565-13571(1989).			
RL	[2]			
RP	SEQUENCE OF 46-1257 FROM N.A.			
RC	TISSUE=Placenta;			
RX	MEDLINE=88083584; PubMed=3691802;			
RA	Soininen R., Haka R., Risku T., Prockop D.J., Tryggvason K.;			
RT	"Complete primary structure of the alpha 1-chain of human basement membrane (type IV) collagen";			
DT	01-FEB-1996 (Rel. 01, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Collagen alpha 1(IV) chain precursor.			
GN	COL4A1.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=8802471; PubMed=3311751;			
RA	Brazel D., Oberbaumer I., Dierlinger H., Babel W., Glanville R.W., Deitzmann R., Kuehn K.;			
RA	"Completion of the amino acid sequence of the alpha 1 chain of human basement membrane collagen (type IV) reveals 21 non-triplet interruptions located within the collagenous domain";			
RT	J. Biol. Chem. 168:529-536(1987).			
RL	Eur. J. Biochem. 168:529-536(1987).			
RN	[4]			
RP	SEQUENCE OF 28-243.			
RX	MEDLINE=86004708; PubMed=4043082;			
RA	Glanville R.W., Qian R.Q., Siebold B., Ristelli J., Kuehn K.;			
RT	"Amino acid sequence of the N-terminal aggregation and cross-linking region (7S domain) of the alpha 1 (IV) chain of human basement membrane collagen";			
RT	Eur. J. Biochem. 152:213-219(1985).			
RN	[5]			
RP	SEQUENCE OF 534-147.			
RT	MEDLINE=8503629; PubMed=6434307;			
RA	Babel W., Glanville R.W.;			
RT	"Structure of human-basement-membrane (type IV) collagen. Complete amino-acid sequence of a 914-residue-long pepsin fragment from the alpha 1(IV) chain";			
RT	Eur. J. Biochem. 143:545-556(1984).			
RL	[6]			
RP	SEQUENCE OF 1256-1669 FROM N.A.			

RX	MEDLINE=95207819; PubMed=2581969;	DR	EMBL; M26548; AAA53098.1; JOINED.
RA	Pihlajaniemi T., Tryggvason K., Myers J.C., Kurkinen M., Lebo R., Cheung M.-C., Prockop D.J., Boyd C.D.;	DR	EMBL; M26549; AAA53098.1; JOINED.
RT	"cDNA clones coding for the pro-alpha1(IV) chain of human type IV procollagen reveal an unusual homology of amino acid sequences in two halves of the carboxyl-terminal domain";	DR	EMBL; M26551; AAA53098.1; JOINED.
RT	J. Biol. Chem. 260:7681-7687(1985).	DR	EMBL; M26552; AAA53098.1; JOINED.
RN	[7]	DR	EMBL; M26553; AAA53098.1; JOINED.
RP	SEQUENCE OF 1259-1669 FROM N.A.	DR	EMBL; M26554; AAA53098.1; JOINED.
RX	MEDLINE=95216555; PubMed=2582422;	DR	EMBL; M26555; AAA53098.1; JOINED.
RA	Brinker J.M., Gudas L.J., Loidl H.R., Wang S.-Y., Rosenblum J., Refalides N.A., Myers J.C.;	DR	EMBL; M26556; AAA53098.1; JOINED.
RA	"Restricted homology between human alpha 1 type IV and other procollagen chains";	DR	EMBL; M26557; AAA53098.1; JOINED.
RT	Proc. Natl. Acad. Sci. U.S.A. 82:3649-3653(1985).	DR	EMBL; M26558; AAA53098.1; JOINED.
RN	[8]	DR	EMBL; M26559; AAA53098.1; JOINED.
RX	SEQUENCE OF 1-28 FROM N.A.	DR	EMBL; M26560; AAA53098.1; JOINED.
RA	Soininen R., Huotari M., Hostikka S.L., Prockop D.J., Tryggvason K.:	DR	EMBL; M26561; AAA53098.1; JOINED.
RT	"The structural genes for alpha 1 and alpha 2 chains of human type IV collagen are divergently encoded on opposite DNA strands and have an overlapping promoter region";	DR	EMBL; M26562; AAA53098.1; JOINED.
RT	J. Biol. Chem. 263:17217-17220(1988).	DR	EMBL; M26563; AAA53098.1; JOINED.
RN	[9]	DR	EMBL; M26564; AAA53098.1; JOINED.
RP	SEQUENCE OF 1441-1669, AND DISULFIDE BONDS.	DR	EMBL; M26565; AAA53098.1; JOINED.
RC	TISSUE=Placenta;	DR	EMBL; M26566; AAA53098.1; JOINED.
RX	MEDLINE=95005112; PubMed=2844531;	DR	EMBL; M26567; AAA53098.1; JOINED.
RA	Siebold B., Deutmann R., Kuhn K.:	DR	EMBL; M26568; AAA53098.1; JOINED.
RT	"The arrangement of intra- and intermolecular disulfide bonds in the carboxyterminal, non-collagenous domain of basement-membrane type IV collagen.",	DR	EMBL; M26569; AAA53098.1; JOINED.
RT	Eur. J. Biochem. 176:617-624(1988).	DR	EMBL; M26570; AAA53098.1; JOINED.
RL	FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF GLOMERULAR BASEMENT MEMBRANES (GBM), FORMING A 'CHICKEN-WIRE' MESHWORK TOGETHER WITH LAMININS, PROTEOGLYCANS AND ENACTIN/ NIDOCIN.	DR	EMBL; M26571; AAA53098.1; JOINED.
CC	-!- SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOFORMS, ALPHA 1(IV)- ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.	DR	EMBL; M26572; AAA53098.1; JOINED.
CC	-!- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS DOMAIN (NC1) AT THEIR C-TERMINUS; FREQUENT INTERRUPTIONS OF THE G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL TRIPLE-HELICAL 7S DOMAIN.	DR	EMBL; M26573; AAA53098.1; JOINED.
CC	-!- PTM: LYSINES AT THE THIRD POSITION OF THE TRIPETIDE REPEATING UNIT ARE HYDROXYLATED IN ALL CASES AND BIND CARBOHYDRATE.	DR	EMBL; M26574; AAA53098.1; JOINED.
CC	-!- DOMAIN: PROLINES AT THE THIRD POSITION OF THE TRIPETIDE REPEATING UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.	DR	EMBL; M26575; AAA53098.1; JOINED.
CC	-!- PTM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE IV COLLAGENS.	DR	EMBL; Y00706; CAA29075.1; -.
CC	-!- PTM: PROLINE AND CYSTEINE RESIDUES WHICH ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE IV COLLAGENS.	DR	EMBL; M10340; AAA52006.1; -.
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to license@sb-sib.ch).	DR	EMBL; M11315; AAA52042.1; -.
CC	-----	DR	PIR; S16876; CSHU4B.
CC	-----	DR	PIR; 121030; -.
CC	-----	DR	InterPro; IPR01442; C4.
CC	-----	DR	InterPro; IPR00087; Collagen.
CC	-----	DR	Pfam; PF01413; C4; 2.
CC	-----	DR	Pfam; PF01391; Collagen; 21.
CC	-----	DR	SMART; P003923; C4; 2.
CC	-----	DR	KW Extracellular matrix; Connective tissue; Basement membrane; Signal Repeat; Hydroxylation; Glycoprotein; Collagen; Signal.
CC	-----	FT	SIGNAL 1 27
CC	-----	FT	PROPEP 28 172
CC	-----	FT	CHAIN 173 1669
CC	-----	FT	DOMAIN 173 1440
CC	-----	FT	DOMAIN 1441 1669
CC	-----	FT	CARBODY 126 126
CC	-----	FT	DISULFID 1460 1551
CC	-----	FT	DISULFID 1493 1548
CC	-----	FT	DISULFID 1505 1511
CC	-----	FT	DISULFID 1570 1665
CC	-----	FT	DISULFID 1604 1662
CC	-----	FT	DISULFID 1616 1622
CC	-----	FT	DISULFID 237 238
CC	-----	FT	DISULFID 241 241
CC	-----	FT	G -> K (IN REF. 4).
CC	-----	FT	O -> A (IN REF. 3).
CC	-----	FT	N -> D (IN REF. 5).
CC	-----	FT	D -> Y (IN REF. 5).
CC	-----	FT	D -> Y (IN REF. 5).
CC	-----	FT	CONFFLICT 837 837
CC	-----	FT	CONFFLICT 842 842
CC	-----	FT	CONFFLICT 895 896
CC	-----	FT	V -> W (IN REF. 2).
CC	-----	FT	E -> Q (IN REF. 5).
CC	-----	FT	S -> K (IN REF. 5).
CC	-----	FT	S -> K (IN REF. 5).
CC	-----	FT	R -> P (IN REF. 5).
CC	-----	FT	S -> K (IN REF. 5).
CC	-----	FT	E -> Q (IN REF. 5).
CC	-----	FT	CONFFLICT 1010 1012
CC	-----	FT	S -> K (IN REF. 5).
CC	-----	FT	CONFFLICT 1012 1012
CC	-----	FT	S -> K (IN REF. 5).
CC	-----	FT	E -> Q (IN REF. 5).
CC	-----	FT	SEQUENCE 1358 1358
CC	-----	FT	1669 AA; 160611 MW.; 3BEB6DDFFB9B8A84 CRC64;
DR	-----	SO	-----
DR	-----	Query Match	100 %; Score 92; DB 1; Length 1669;
DR	-----	Best Local Similarity	100.0%; Pred. No. 5.3e-06;

Matches 15; conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 2

Oy 1 GVKGDKGNGPGWCGAP 15  
 ID CA14\_MOUSE STANDARD; PRT; 1669 AA.  
 AC P02463;  
 DT 01-FEB-1991 (Rel. 17, last sequence update)  
 DT 21-JUL-1986 (Rel. 01, created)  
 DE Collagen alpha 1(IV) chain precursor.  
 GN COL4A1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 RN NEBI\_TAXID=10090;  
 RP [1]  
 SEQUENCE FROM N.A.  
 MEDLINE=89197932; PubMed=2703490;  
 RA Muthukumaran G., Blumberg B., Kurkinen M.;  
 RT "the complete primary structure for the alpha 1-chain of mouse collagen IV. Differential evolution of collagen IV domains.";  
 RL J. Biol. Chem. 264:6310-6317(1989).  
 RN [2]  
 SEQUENCE OF 1-1154 FROM N.A.  
 MEDLINE=88112221; PubMed=3338568;  
 RX Wood L., Theriault N., Vogeli G.;  
 RT cDNA clones completing the nucleotide and derived amino acid sequence of the alpha 1 chain of basement membrane (type IV) collagen from mouse.";  
 RL FEBS Lett. 227:5-8(1988).  
 RN [3]  
 SEQUENCE OF 1149-1424 FROM N.A.  
 MEDLINE=86301886; PubMed=3755692;  
 RA Nath P., Laurent M., Horn E., Sobel M.E., Zon G., Vogeli G.;  
 RT "isolation of an alpha 1 type-IV collagen cDNA clone using a synthetic oligodeoxynucleotide.";  
 RL Gene 43:301-304(1986).  
 RN [4]  
 RP SEQUENCE OF 1276-1669 FROM N.A.  
 RX MEDLINE=8512033; PubMed=2379861;  
 RA Oberbaumer T., Laurent M., Schwarz U., Sakurai Y., Yamada Y.,  
 RA Vogeli G., Voss T., Siebold B., Glanville R.W., Kuhn K.;  
 RT "Amino acid sequence of the non-collagenous globular domain (NC1) of the alpha 1(IV) chain of basement membrane collagen as derived from complementary DNA.";  
 RL Eur. J. Biochem. 147:217-224(1985).  
 RN [5]  
 SEQUENCE OF 1441-1669 FROM N.A.  
 MEDLINE=87230460; PubMed=3597383;  
 RA Kurkinen M., Condon M.R., Blumberg B., Barlow D., Quinones S.,  
 RA Saus J., Phlajanciemi T.;  
 RT "Extensive homology between the carboxyl-terminal peptides of mouse alpha 1(IV) and alpha 2(IV) collagen.";  
 RL J. Biol. Chem. 262:8496-8499(1987).  
 RN [6]  
 RP PARTIAL SEQUENCE FROM N.A.  
 RX MEDLINE=86156099; PubMed=3009468;  
 RA Sakurai Y., Sullivan M., Yamada Y.;  
 RT "Alpha 1 type IV collagen gene evolved differently from fibrillar collagen genes.";  
 RL J. Biol. Chem. 261:6654-6657(1986).  
 RN [7]  
 RP SEQUENCE OF 1-28 FROM N.A.  
 RX MEDLINE=89056738; PubMed=3198626;  
 RA Keytes P., Wood L., Theriault N., Kurkinen M., Vogeli G.;  
 RT "Head-to-head arrangement of murine type IV collagen genes.";  
 RL J. Biol. Chem. 263:19274-19277(1988).  
 RN [8]

RP SEQUENCE OF 1-28 FROM N.A.  
 RX MEDLINE=89071759; PubMed=3200851;  
 RA Burbelo P.D., Martin G.R., Yamada Y.;  
 RT "Alpha 1(IV) and alpha 2(IV) collagen genes are regulated by a bidirectional promoter and a shared enhancer.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 85:9579-9582(1988);  
 RN [9]

RP SEQUENCE OF 1-129 FROM N.A.  
 RX MEDLINE=88243724; PubMed=3379041;  
 RA Killen P.D., Burbelo P., Sakurai Y., Yamada Y.;  
 RT "Structure of the amino-terminal portion of the murine alpha 1(IV) collagen chain and the corresponding region of the gene.";  
 RL J. Biol. Chem. 263:8706-8709(1988).  
 CC -1- FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF GLOMERULAR BASEMENT MEMBRANES (GBM), FORMING A 'CHICKEN-WIRE' MESHWORK TOGETHER WITH LAMININS, PROTEOGLYCANS AND ENACTIN/ NIDOCIN.

CC -1- SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOFORMS, ALPHA 1(IV)-ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.

CC -1- DOMAIN: CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL TRIPLE-HELICAL 7S DOMAIN.

CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPETIDE REPEATING UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.

CC -1- TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH ARE INVOLVED IN INNER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE IV COLLAGENS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC ---

DR EMBL; J03758; AAA37439.1; ---  
 DR EMBL; M23333; AAA51625.1; ---  
 DR EMBL; J04694; AAA50292.1; ---  
 DR EMBL; X06777; CAM29946.1; ---  
 DR EMBL; X02201; CAM26132.1; ---  
 DR EMBL; M15832; AAA37340.1; ---  
 DR EMBL; M14042; AAA37342.1; ---  
 DR EMBL; M12879; AAA37343.1; ---  
 DR EMBL; M13024; ---; NOT\_ANNOTATED\_CDS.  
 DR EMBL; M13025; ---; NOT\_ANNOTATED\_CDS.  
 DR EMBL; M13026; AAA37344.1; ---  
 DR EMBL; M13027; AAA37345.1; ---  
 DR EMBL; M13043; AAA37346.1; ---  
 DR EMBL; J04448; AAA37437.1; ---  
 DR PIR; A33525; CGMSB.  
 DR MGD; MGI:88454; Col4al.  
 DR InterPro; IPR00142; C4.  
 DR InterPro; IPR00087; Collagen.  
 DR Pfam; PF01413; C4; 2.  
 DR Pfam; PF01391; Collagen; 21.  
 DR Prodrom; PDD03923; C4; 2.  
 DR SMART; SM00111; C4; 2.  
 KW Extracellular matrix; Connective tissue; Basement membrane; Repeat; Hydroxylation; Signal.

FT SIGNAL 1 27  
 FT PROPEP 28 172 AMINO-TERMINAL PROPEPTIDE (7S DOMAIN).

FT CHAIN 173 1669 COLLAGEN ALPHA 1(IV) CHAIN.  
 FT DOMAIN 173 1440 TRIPLE-HELICAL REGION.  
 FT DOMAIN 1441 1669 NONHELICAL REGION (NC1).

FT DISULFID 1460 1551 OR 1548 (BY SIMILARITY).  
 FT DISULFID 1493 1548 OR 1551 (BY SIMILARITY).  
 FT DISULFID 1505 1511 BY SIMILARITY.

FT DISULFID 1570 1665 OR 1662 (BY SIMILARITY).

Query Match 84 8%; Score 78; DB 1; Length 1669;  
 Best Local Similarity 86.7%; Pred. No. 0.00065; Matches 13; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GVKGDKGNPGWPGAP 15  
 Db 1263 GPKGDKGNGQWPGAP 1277

RESULT 3

CAS5\_HUMAN STANDARD; PRT: 1685 AA.

ID P29400; 016126; 01006; AC  
 DT 01-DEC-1992 (Rel. 24, Created)  
 DT 01-FEB-1994 (Rel. 28, last sequence update)  
 DT 01-MAR-2002 (Rel. 41, last annotation update)

DE Collagen alpha 5(IV) chain precursor.  
 GN COL4A5.

OS Homo sapiens (Human).

OC Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OX Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

RN [1] NCBI\_TaxID=9606;

RN SEQUENCE FROM N.A.

RX MEDLINE=94165049; PubMed=8120014;  
 RA Zhou J., Leinonen A., Tryggvason K.;  
 RT "Structure of the human type IV collagen COL4A5 gene.";  
 RL J. Biol. Chem. 269:6608-6614(1994).  
 [2]

RN SEQUENCE OF 1-910 FROM N.A., AND VARIANT AS CYS-521.

RX TISSUE=Kidney;  
 MEDLINE=92315623; PubMed=1352287;  
 RA Zhou J., Hertz J.M., Leinonen A., Tryggvason K.;  
 RT "Complete amino acid sequence of the human alpha 5 (IV) collagen  
 chain and identification of a single-base mutation in exon 23  
 converting glycine 521 in the collagenous domain to cysteine in an  
 Alport syndrome patient.";  
 RL J. Biol. Chem. 267:12475-12481(1992).  
 [3]

RN SEQUENCE OF 85-1685 FROM N.A.

RX TISSUE=Placenta;  
 MEDLINE=00337950; PubMed=2380186;

RA Pihlajaniemi T., Pohjolainen E.R., Myers J.C.;  
 RT "Complete primary structure of the triple-helical region and the  
 carboxyl-terminal domain of a new type IV collagen chain, alpha  
 5(IV)." ;  
 J. Biol. Chem. 265:13758-13766(1990).

RN [4] SEQUENCE OF 924-1685 FROM N.A.

RX MEDLINE=91169491; PubMed=2004755;  
 RA Zhou J., Hostikka S.L., Chow L.T., Tryggvason K.;  
 RT "Characterization of the half of the human type IV collagen alpha  
 5 gene that is affected in the Alport syndrome.";  
 RL Genomics 9:1-9(1991).  
 [5]

RX SEQUENCE OF 914-1685 FROM N.A.

RX MEDLINE=90160375; PubMed=1689491;

RA Hostikka S.L., Eddy R.L., Byers M.G., Hoeyhvae M., Shows T.B., Tryggvason K.;  
 RA "Identification of a distinct type IV collagen alpha chain with restricted kidney distribution and assignment of its gene to the locus of X chromosome-linked Alport syndrome.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 87:1606-1610(1990).  
 RN [6] SEQUENCE OF 1442-1471 FROM N.A.

RX MEDLINE=90252791; PubMed=2333699;

RA Myers J.C., Jones T.A., Pohjolainen E.R., Kadri A.S., Goddard A.D., Sheer D., Solomon E., Pihlajaniemi T.;  
 RT "Molecular cloning of the X chromosome containing the gene to the region of the X chromosome containing the Alport syndrome locus.";  
 RL An. J. Hum. Genet. 46:1024-1033(1990).  
 RN [7] SEQUENCE OF 1-20 FROM N.A.

RA Guo C., van Damme B., Vanrenterghem Y., Devriendt K., Cassiman J.-J., Marynen P.;  
 RL Submitted (SEP-1994) to the EMBL/GenBank/DDJB databases.  
 RN [8] SEQUENCE OF 1258-1270 FROM N.A. (SPLICED FORM).

RX MEDLINE=94133540; PubMed=8301933;

RA Guo C., van Damme B., van den Berghe H., Cassiman J.-J., Marynen P.;  
 RT "Differential splicing of COL4A5 mRNA in kidney and white blood cells: a complex mutation in the COL4A5 gene of an Alport patient deletes the NC1 domain.";  
 RL Kidney Int. 44:1316-1321(1993).  
 RN [9]

RX MEDLINE=97338662; PubMed=9195222;

RA Lemmink H.H., Schroeder C.H., Monnens L.A.H., Smeets H.J.M.;  
 RT "The clinical spectrum of type IV collagen mutations.";  
 RL Hum. Mutat. 9:477-499(1997).  
 RN [10]

RX MEDLINE=91169492; PubMed=1672282;

RA Zhou J., Barker D.F., Hostikka S.L., Gregory M.C., Atkins C.L., Tryggvason K.;  
 RT "Single base mutation in alpha 5(IV) collagen chain gene converting a conserved cysteine to serine in Alport syndrome.";  
 RL Genomics 9:10-18(1991).  
 RN [11]

RX MEDLINE=95303539; PubMed=1376965;

RA Knebelmann B., Deschenes G., Gros F., Hors M.-C., Gruenfeld J.-P., Tryggvason K., Gubler M.-C., Antignac C.;  
 RT "Substitution of arginine for glycine 325 in the collagen alpha 5 (IV) chain associated with X-linked Alport syndrome: characterization of the mutation by direct sequencing of PCR-amplified lymphoblast DNA fragments.";  
 RT Ann. J. Hum. Genet. 51:135-142(1992).  
 RL [12]

RX MEDLINE=93244772; PubMed=1363780;

RA Renieri A., Sarri M., Myers J.C., Pihlajaniemi T., Massella L., Rizzoni G.F., de Marchi M.;  
 RT "De novo mutation in the COL4A5 gene converting glycine 325 to glutamic acid in Alport syndrome.";  
 RT Hum. Mol. Genet. 1:127-129(1992).  
 RL [13]

RX MEDLINE=9401048; PubMed=8400498;

RA Lemmink H.L., Schroeder C.H., Brunner H.G., Nelen M.R., Zhou J., Tryggvason K., Haggema-Schouten W.A.G., Roordaets A.P., Rascher W., van Oost B.A., Smeets H.J.M.;  
 RT "Identification of four novel mutations in the COL4A5 gene of patients with Alport syndrome.";  
 RL Genomics 17:485-489(1993).  
 RN [14]

RX MEDLINE=90160375; PubMed=1689491;

RA Variants AS E-400; V-406; V-638; A-638; R-653; R-796; R-869; R-872  
 AND C-1241.

- RX MEDLINE=95322976; PubMed=7599631;  
 RA Bove E., Flinter F., Zhou J., Tryggvason K., Borrows M., Harris A.;  
 RT "detection of 12 novel mutations in the collagenous domain of the  
 COL4A5 gene in Alport syndrome patients.";  
 RT Hum. Mutat. 5:197-204(1995).  
 RN [15]  
 RP VARIANT AS ARG-1649.  
 RX MEDLINE=96213750; PubMed=86551292;  
 RA Barker D.F., Pruchino C.J., Jiang X., Atkin C.L., Stone E.M.,  
 Denison J.C., Fair P.R., Gregory M.C.;  
 RT "A mutation causing Alport syndrome with tardive hearing loss is  
 common in the western United States";  
 RL Am. J. Hum. Genet. 58:1157-1165(1996).  
 RN [16]  
 RP VARIANT AS.  
 RX MEDLINE=96213754; PubMed=86551296;  
 RA Renieri A., Bruttini M., Galli L., Zarelli P., Neri T.M., Rossetti S.,  
 RA Scalari F., Sesia A., Rizzoni G.F., Tryggvason K., Pignatti P.F.,  
 RA Savi M., Ballabio A., de Marchi M.;  
 RT "X-linked Alport syndrome: an SSPC-based mutation survey over all 51  
 exons of the COL4A5 gene.";  
 RL Am. J. Hum. Genet. 58:1192-1204(1996).  
 RN [17]  
 RP VARIANT AS, AND VARIANT AS-P-430; SER-444; SER-619; ASN-664 AND  
 RP MEDLINE=97094179; PubMed=8940267;  
 RA Knebelmann B., Breillat C., Forestier L., Arrendel C., Jaccassier D.,  
 RA Gistras T., Drouot L., Deschenes G., Grunfeld J.-P., Broyer M.,  
 RA Gabler M.-C., Antignac C.;  
 RT "Spectrum of mutations in the COL4A5 collagen gene in X-linked Alport  
 syndrome.,";  
 RL Am. J. Hum. Genet. 59:1221-1232(1996).  
 RN [18]  
 RP VARIANT AS ASP-198.  
 RX MEDLINE=96223392; PubMed=8829632;  
 RA Tverskaya S., Bobryntseva V., Tsalykova F., Ignatova M.,  
 RA Krasnopol'skaya X., Evrakov O.;  
 RT "Common ancestry of three Ashkenazi-American families with Alport  
 chain associated with adult-onset X-linked Alport syndrome.";  
 RL Hum. Mutat. 7:149-150(1996).  
 RN [19]  
 RP VARIANT AS GLN-167.  
 RX MEDLINE=97295089; PubMed=9150741;  
 RA Barker D.F., Denison J.C., Atkin C.L., Gregory M.C.;  
 RT "Common ancestry of three Ashkenazi-American families with Alport  
 syndrome and COL4A5 R167Q.";  
 RL Hum. Genet. 99:681-684(1997).  
 RN [20]  
 RP VARIANT AS R-174; R-177; R-325; C-1410; W-1421; T-1517 AND D-1596.  
 RX MEDLINE=98124435; PubMed=9152056;  
 RA Neri T.M., Zanelli P., de Palma G., Savi M., Rossetti S., Turco A.E.,  
 RA Trivellati G.F., Galli L., Bruttini M., Renieri A., Mangarelli R.,  
 RA Trivellati A., Pinciaroli A.R., Reggiani M., Rizzoni G.F., de Marchi M.;  
 RT "Missense mutations in the COL4A5 gene in patients with X-linked  
 Alport syndrome.,";  
 RL Hum. Mutat. Suppl. 1:S106-S109(1998).  
 RN [21]  
 RP VARIANT AS V-420; 456-P--P-458 DEL; D-573; D-624; D-635; 802-G--P-807  
 DEL; C-941; S-1030; S-1066; D-1143; R-1196; E-1261; S-1357  
 RP AND R-1649.  
 RX MEDLINE=99063529; PubMed=98488783;  
 RA Martin P., Heiskanen N., Zhou J., Leinonen A., Tumelius T., Hertz J.M.,  
 RA Barker D.F., Gregory M.C., Atkin C.L., Styrkardottir U., Neumann H.,  
 RA Springer J., Shows T.B., Pettersson E., Tryggvason K.;  
 RT "High mutation detection rate in the COL4A5 collagen gene in suspected  
 RT Alport syndrome using PCR and direct DNA sequencing.,";  
 RL J. Am. Soc. Nephrol. 9:2291-2301(1998).  
 RN [22]  
 RP VARIANT AS GLU-579; LYS-623; ASN-947; VAL-953; ARG-1107; ARG-1158;  
 RP SER-1170 AND TRP-1167, AND VARIANT SER-444 AND ALA-739.  
 RX MEDLINE=20030197; PubMed=10561141;  
 RA Inoue Y., Nishio H., Shirakawa T., Nakaniishi K., Nakamura H.,

RX MEDLINE=20025011; PubMed=10563487;  
 RX Query Match 75.0%; Score 69; DB 1; Length 1685;  
 RT Best Local Similarity 73.3%; Pred. No. 0.014;  
 RT Matches 11; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
 RN QY 1 GVKGDKGNRGWPGAP 15  
 Db 1279 GIKEKGKGNFGQPSLP 1293

RESULT 4  
 CAS4\_CANFA STANDARD: PRY: 754 AA.  
 ID CAS4\_CANFA  
 AC Q2247;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE Collagen alpha 5(IV) chain (Fragment).  
 GN COL4A5.  
 OS Canis familiaris (Dog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi;  
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
 OX NCBI\_TAXID=9615;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-SAMOTED; TISSUE-Kidney;  
 RX MEDLINE=94224868; PubMed=8171024;  
 RA Zheng K., Thornton P.S., Marrano P., Baumal R., McInnes R.R.;  
 RT "Canine X chromosome-linked hereditary nephritis: a genetic model for  
 human X-linked hereditary nephritis resulting from a single base  
 mutation in the gene encoding the alpha 5 chain of collagen type  
 IV.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 91:3989-3993(1994).  
 RL -1- FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF  
 CC GLOMERULAR BASEMENT MEMBRANES (GBM), FORMING A 'CHICKEN-WIRE'  
 CC MESHWORK TOGETHER WITH LAMININS, PROTEOGLYCANS AND ENACTIN/  
 CC NIDGEN.  
 CC -1- SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOFORMS, ALPHA 1(IV)-  
 CC ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE  
 CC WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.  
 CC -1- SUBCELLULAR LOCATION: CELL SURFACE (POTENTIAL).  
 CC -1- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS  
 CC DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE  
 CC G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY  
 CC CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL  
 CC TRIPLE-HELICAL 7S DOMAIN.  
 CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPETIDE REPEATING  
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
 CC -1- DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE  
 CC G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY  
 CC CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL  
 CC TRIPLE-HELICAL 7S DOMAIN.  
 CC -1- DISEASE: A DEFECT IN COL4A5 HAS BEEN FOUND TO BE THE CAUSE OF  
 CC CANINE X-LINKED HEREDITARY NEPHRITIS (HN), A DISEASE SIMILAR TO  
 CC THAT IN HUMANS (ALSO REFERRED TO AS ALPORT SYNDROME) CHARACTERIZED  
 CC BY PROGRESSIVE RENAL FAILURE AND NEUROSENSORY DEAFNESS.  
 CC -1- SIMILARITY: TO OTHER TYPE IV COLLAGENS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)  
 CC or send an email to license@isb-sib.ch).



DR InterPro; IPR01442; C4  
 DR InterPro; IPR000087; Collagen.  
 DR Pfam; PF01413; C4; 2.  
 DR Pfam; PF01391; Collagen; 19.  
 DR PRODOM; PD00323; C4; 2.  
 DR SMART; SM00111; C4; 2.  
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
 Glycoprotein; Basement membrane; Collagen; Signal.  
 FT SIGNAL 1 25  
 FT PROPEP 26 183  
 FT CHAIN 184 1707  
 FT DOMAIN 184 1479  
 FT DOMAIN 1480 1707  
 FT DISULFID 1499 1588  
 FT DISULFID 1532 1585  
 FT DISULFID 1544 1550  
 FT DISULFID 1607 1703  
 FT DISULFID 1641 1700  
 FT DISULFID 1653 1660  
 FT CARBOHYD 138 138  
 FT CONFLICT 1051 1051  
 FT CONFLICT 1097 1171  
 FT CONFLICT 1171 1171  
 FT CONFLICT 1179 1179  
 FT CONFLICT 1241 1241  
 FT CONFLICT 1328 1328  
 FT CONFLICT 1573 1573  
 FT CONFLICT 1623 1623  
 SEQENCE 1670 AA; 167391 MW; 1A56159605FD508 CRC64;  
 Query Match 70.7%; Score 65; DB 1; Length 1707;  
 Best Local Similarity 66.67%; Pred. No. 058; RT  
 Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 GVKGDKGNPGWPGV 15  
 Db 1165 GIPGDKGDFGWPVGAP 15

**RESULT 6**

**CA24\_HUMAN STANDARD;** PRT: 1712 AA.

ID CA24\_HUMAN STANDARD; PRT: 1712 AA.

AC P08572; DT 01-AUG-1988 (Rel. 08, Created)  
 DT 01-FEB-1991 (Rel. 17, Last sequence update)  
 DT 2001-10-20 (Rel. 40, last annotation update)

DE Collagen alpha 2(IV) chain precursor.

GN COLA2.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TAXID=9606;

RN [1] SEQUENCE FROM N.A. MEDLINE=99066769; PubMed=3198637;

RX RA Hostikka S.L., Tryggvason K.; "The complete primary structure of the alpha 2 chain of human type IV collagen and comparison with the alpha 1(IV) chain.", J. Biol. Chem. 263:19488-19503(1988).

RN [2] SEQUENCE OF 1-1042 FROM N.A. MEDLINE=8815998; PubMed=3345760;

RX RA Brazel D., Pollner R., Oberhaeuser I., Kuehn K.; "Human basement membrane collagen (type IV). The amino acid sequence of the alpha 2(IV) chain and its comparison with the alpha 1(IV) chain reveals deletions in the alpha 1(IV) chain.", Eur. J. Biochem. 172:35-42(1988).

RN [3] SEQUENCE OF 1254-1712 FROM N.A. MEDLINE=8721958; PubMed=3582677;

RX RA Hostikka S.L., Kurkinen M., Tryggvason K.;

**RESULT 6**

**CA24\_HUMAN STANDARD;** PRT: 1712 AA.

ID CA24\_HUMAN STANDARD; PRT: 1712 AA.

AC P08572; DT 01-AUG-1988 (Rel. 08, Created)  
 DT 01-FEB-1991 (Rel. 17, Last sequence update)  
 DT 2001-10-20 (Rel. 40, last annotation update)

DE Collagen alpha 2(IV) chain precursor.

GN COLA2.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TAXID=9606;

RN [1] SEQUENCE FROM N.A. MEDLINE=99066769; PubMed=3198637;

RX RA Hostikka S.L., Tryggvason K.; "The complete primary structure of the alpha 2 chain of human type IV collagen and comparison with the alpha 1(IV) chain.", J. Biol. Chem. 263:19488-19503(1988).

RN [2] SEQUENCE OF 1-1042 FROM N.A. MEDLINE=8815998; PubMed=3345760;

RX RA Brazel D., Pollner R., Oberhaeuser I., Kuehn K.; "Human basement membrane collagen (type IV). The amino acid sequence of the alpha 2(IV) chain and its comparison with the alpha 1(IV) chain reveals deletions in the alpha 1(IV) chain.", Eur. J. Biochem. 172:35-42(1988).

RN [3] SEQUENCE OF 1254-1712 FROM N.A. MEDLINE=8721958; PubMed=3582677;

RX RA Hostikka S.L., Kurkinen M., Tryggvason K.;

**AMINO-TERMINAL PROPEPTIDE (7S DOMAIN).**

FT SIGNAL 1 25  
 FT PROPEP 26 183  
 FT CHAIN 184 1707  
 FT DOMAIN 184 1479  
 FT DOMAIN 1480 1707  
 FT DISULFID 1499 1588  
 FT DISULFID 1532 1585  
 FT DISULFID 1544 1550  
 FT DISULFID 1607 1703  
 FT DISULFID 1641 1700  
 FT DISULFID 1653 1660  
 FT CARBOHYD 138 138  
 FT CONFLICT 1051 1051  
 FT CONFLICT 1097 1171  
 FT CONFLICT 1171 1171  
 FT CONFLICT 1179 1179  
 FT CONFLICT 1241 1241  
 FT CONFLICT 1328 1328  
 FT CONFLICT 1573 1573  
 FT CONFLICT 1623 1623  
 SEQENCE 1670 AA; 167391 MW; 1A56159605FD508 CRC64;

Query Match 70.7%; Score 65; DB 1; Length 1707;  
 Best Local Similarity 66.67%; Pred. No. 058; RT  
 Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 GVKGDKGNPGWPGV 15  
 Db 1165 GIPGDKGDFGWPVGAP 15

**SEQUENCE OF 1451-1485 FROM N.A.**

FT SIGNAL 1 25  
 FT PROPEP 26 183  
 FT CHAIN 184 1707  
 FT DOMAIN 184 1479  
 FT DOMAIN 1480 1707  
 FT DISULFID 1499 1588  
 FT DISULFID 1532 1585  
 FT DISULFID 1544 1550  
 FT DISULFID 1607 1703  
 FT DISULFID 1641 1700  
 FT DISULFID 1653 1660  
 FT CARBOHYD 138 138  
 FT CONFLICT 1051 1051  
 FT CONFLICT 1097 1171  
 FT CONFLICT 1171 1171  
 FT CONFLICT 1179 1179  
 FT CONFLICT 1241 1241  
 FT CONFLICT 1328 1328  
 FT CONFLICT 1573 1573  
 FT CONFLICT 1623 1623  
 SEQENCE 1670 AA; 167391 MW; 1A56159605FD508 CRC64;

Query Match 70.7%; Score 65; DB 1; Length 1707;  
 Best Local Similarity 66.67%; Pred. No. 058; RT  
 Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 GVKGDKGNPGWPGV 15  
 Db 1165 GIPGDKGDFGWPVGAP 15

**SEQUENCE OF 1886-1712 FROM N.A.**

FT SIGNAL 1 25  
 FT PROPEP 26 183  
 FT CHAIN 184 1707  
 FT DOMAIN 184 1479  
 FT DOMAIN 1480 1707  
 FT DISULFID 1499 1588  
 FT DISULFID 1532 1585  
 FT DISULFID 1544 1550  
 FT DISULFID 1607 1703  
 FT DISULFID 1641 1700  
 FT DISULFID 1653 1660  
 FT CARBOHYD 138 138  
 FT CONFLICT 1051 1051  
 FT CONFLICT 1097 1171  
 FT CONFLICT 1171 1171  
 FT CONFLICT 1179 1179  
 FT CONFLICT 1241 1241  
 FT CONFLICT 1328 1328  
 FT CONFLICT 1573 1573  
 FT CONFLICT 1623 1623  
 SEQENCE 1670 AA; 167391 MW; 1A56159605FD508 CRC64;

Query Match 70.7%; Score 65; DB 1; Length 1707;  
 Best Local Similarity 66.67%; Pred. No. 058; RT  
 Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 GVKGDKGNPGWPGV 15  
 Db 1165 GIPGDKGDFGWPVGAP 15

**SEQUENCE OF 1-33 FROM N.A.**

FT SIGNAL 1 25  
 FT PROPEP 26 183  
 FT CHAIN 184 1707  
 FT DOMAIN 184 1479  
 FT DOMAIN 1480 1707  
 FT DISULFID 1499 1588  
 FT DISULFID 1532 1585  
 FT DISULFID 1544 1550  
 FT DISULFID 1607 1703  
 FT DISULFID 1641 1700  
 FT DISULFID 1653 1660  
 FT CARBOHYD 138 138  
 FT CONFLICT 1051 1051  
 FT CONFLICT 1097 1171  
 FT CONFLICT 1171 1171  
 FT CONFLICT 1179 1179  
 FT CONFLICT 1241 1241  
 FT CONFLICT 1328 1328  
 FT CONFLICT 1573 1573  
 FT CONFLICT 1623 1623  
 SEQENCE 1670 AA; 167391 MW; 1A56159605FD508 CRC64;

Query Match 70.7%; Score 65; DB 1; Length 1707;  
 Best Local Similarity 66.67%; Pred. No. 058; RT  
 Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 GVKGDKGNPGWPGV 15  
 Db 1165 GIPGDKGDFGWPVGAP 15

**"Nucleotide sequence coding for the human type IV collagen alpha 2 chain cDNA reveals extensive homology with the NC-1 domain of alpha 1 (IV) but not with the collagenous domain or 3'-untranslated region."**

RT Griffin C.A., Emanuel B.S., Hansen J.R., Cavenee W.K., Myers J.C.; "Human collagen genes encoding basement membrane alpha 1 (IV) and alpha 2 (IV) chains map to the distal long arm of chromosome 13."; FEBS Lett. 216:281-286(1987).

RL [4] [5]

RN [6]

RP SEQUENCE OF 1451-1485 FROM N.A.

RX MEDLINE=87092438; PubMed=3025878;

RA Griffin C.A., Emanuel B.S., Hansen J.R., Cavenee W.K., Myers J.C.; "Human collagen genes encoding basement membrane alpha 1 (IV) and alpha 2 (IV) chains map to the distal long arm of chromosome 13."; FEBS Lett. 216:281-286(1987).

RT Proc. Natl. Acad. Sci. U.S.A. 84:512-516(1987).

RL [7]

RP SEQUENCE OF 1886-1712 FROM N.A.

RX MEDLINE=87250511; PubMed=2439508;

RA Soiminen R., Hiotari M., Hostikka S.L., Prockop D.J., Tryggvason K.; "The structural genes for alpha 1 and alpha 2 chains of human type IV collagen are divergently encoded on opposite DNA strands and have an overlapping promoter region."; J. Biol. Chem. 263:17217-17220(1988).

RL [8]

RP SEQUENCE OF 1-33 FROM N.A.

RX MEDLINE=88030332; PubMed=2846280;

RA Poeschl E., Pollner R., Kuehn K.; "The genes for the alpha 1(IV) and alpha 2(IV) chains of human basement membrane collagen type IV are arranged head-to-head and separated by a bidirectional promoter of unique structure."; EMBO J. 7:2687-2695(1988).

RL [9]

RP SEQUENCE OF 1-33 FROM N.A.

RX MEDLINE=93305049; PubMed=8317999;

RA Fischer G., Schmidt C., Opitz J., Cully Z., Kuehn K., Poeschl E.; "Identification of a novel sequence element in the common promoter region of human collagen type IV genes, involved in the regulation of divergent transcription"; Bloch. J. 292:687-695(1993).

RL [9]

RP SEQUENCE OF 1880-1535; 1545-1614; 1617-1701 AND 1705-1712.

RC TISSUE=Placenta;

RX MEDLINE=88005112; PubMed=2844531;

RA Siebold B., Deutzmann R., Kuehn K.; "The arrangement of intra- and intermolecular disulfide bonds in the carboxyterminal, non-collagenous aggregation and cross-linking domain of basement-membrane type IV collagen."; Eur. J. Bloch. 176:617-624(1988).

CC -I- FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF GLOMULAR BASEMENT MEMBRANES (GBM), FORMING A 'CHICKEN-WIRE' MESHWORK TOGETHER WITH LAMININS, PROTEOGLYCANS AND ENTACTIN/ NIDGEN.

CC -I- SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOFORMS, ALPHA 1(IV)- ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.

CC -I- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL TRIPLE-HELICAL 7S DOMAIN.

CC -I- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPETIDE REPEATING UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.

CC -I- PTM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE IV COLLAGENS.

CC ---

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation.

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.ebi.ac.uk/announce/> or send an email to licensee@ebi.ac.uk).

CC EMBL; X05562; CAA29076.1; -  
 CC EMBL; X05610; CAA2908.1; --  
 CC EMBL; J02760; AAA58422.1; --  
 CC EMBL; M35953; AAA53099.1; --  
 CC EMBL; X12784; CAA31275.1; --  
 CC EMBL; J04217; AAA53097.1; --  
 CC PIR; A332024; A32024.  
 DR MIM; 120090; -.  
 DR InterPro; IPR01442; C4.  
 DR InterPro; IPR00087; Collagen.  
 DR Pfam; PF01391; Collagen; 20.  
 DR ProdDom; PD00393; C4; 2.  
 DR SMART; SM0011; C4; 2.  
 KW Glycoprotein; Basement membrane; Collagen; Signal.  
 FT SIGNAL 1 25  
 FT PROPEP 26 183  
 FT CHAIN 184 1712  
 FT DOMAIN 1484 1712  
 FT DISULFID 1504 1593  
 FT DISULFID 1537 1590  
 FT DISULFID 1549 1555  
 FT DISULFID 1612 1708  
 FT DISULFID 1646 1705  
 FT DISULFID 1658 1665  
 FT CARBONYD 138 138  
 FT CONFLICT 471 471  
 FT CONFLICT 683 683  
 FT CONFLICT 1575 1575  
 FT CONFLICT 1663 1663  
 FT CONFLICT 1701 1701  
 SQ SEQUENCE 1712 AA; 167535 MW; 2552A1/847890037 CRC64;

Query Match 69.6%; Score 64; DB 1; Length 1712;  
 Best Local Similarity 73.3%; Pred. No. 0.081; Oy 1 GVKGDKGKGNPGMPGAP 15  
 Matches 11; Conservative 2; Mismatches 2; Indels 0; Gaps 0; Sq 1368 GPKGPKGDPGPGAP 1382

RESULT 7

CA17\_HUMAN  
 ID CA17\_HUMAN STANDARD; PRT; 2944 AA.  
 AC Q02388; Q14054; Q16507;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Collagen alpha 1(VII) chain precursor (long-chain collagen) (LC  
 DE collagen).  
 CO LA1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarhini; Hominidae; Homo.  
 OC NCBI\_TAXID=9606;  
 RN [1]  
 RP SEQUENCE FROM N\_A.  
 RX MEDLINE=94327588; PubMed=8051117;  
 RA Christiano A.M., Greenspan D.S., Lee S., Ututto J.;  
 RT "Cloning of human type VII collagen. Complete primary sequence of the alpha 1(VII) chain and identification of intragenic polymorphisms.";  
 J. Biol. Chem. 269:20256-20262(1994).  
 RN [2]  
 RP SEQUENCE OF 128-1493 FROM N\_A., AND PARTIAL SEQUENCE.

RA MEDLINE=93338437; PubMed=1307247;  
 RA Christiano A.M., Rosenbaum L.M., Chung-Honet L.C., Parente M.G., Woodley D.T., Pan C.C., Zhang R.Z., Chu M.-L., Burgesson R.E., Ututto J.; "The large non-collagenous domain (NC-1) of type VII collagen is amino-terminal and chimeric. Homology to cartilage matrix protein, the type III domains of fibronectin and the A domains of von Willebrand factor.;"  
 RL Hum. Mol. Genet. 1:475-481(1992).  
 RN [3]  
 RT SEQUENCE OF 815-1439 FROM N\_A.  
 RA MEDLINE=9134380; Published=1971109;  
 RA Parente M.G., Chung L.C., Rynneen J., Woodley D.T., Wynn K.W., Bauer E.A., Mattel M.-G., Chu M.-L., Ututto J.; "Human type VII collagen: cDNA cloning and chromosomal mapping of the gene.";  
 Proc. Natl. Acad. Sci. U.S.A. 88:6931-6935(1991).  
 RN [4]  
 RP SEQUENCE OF 369-1255 FROM N\_A.  
 RX MEDLINE=93107742; Published=1469284;  
 RA Gammon W.R., Abernethy M.L., Padilla K.M., Prisayanh P.S., Cook M.E., Wright J., Briggaman R.A., Hunt S.W. III;  
 RT "Noncollagenous (NC1) domain of collagen VII resembles multidomain adhesion proteins involved in tissue-specific organization of extracellular matrix.";  
 RT J. Invest. Dermatol. 99: 691-696(1992).  
 RN [5]  
 RP SEQUENCE OF 340-675 FROM N\_A.  
 RC TISSUE-Keratinocytes;  
 RX MEDLINE=9231902; Published=1567409;  
 RA Tanaka T., Takahashi K., Furukawa F., Imamura S.; "Molecular cloning and characterization of type VII collagen cDNA.";  
 RL Biochem. Biophys. Res. Commun. 183:958-963(1992).  
 RN [6]  
 RP SEQUENCE OF 271985-2944 FROM N\_A.  
 RX MEDLINE=93271985; Published=849916;  
 RA Greenspan D.S.; "The carboxyl-terminal half of type VII collagen, including the non-collagenous NC-2 domain and intron/exon organization of the corresponding region of the COL7A1 gene.";  
 RT Hum. Mol. Genet. 2:273-278(1993).  
 RN [7]  
 RP SEQUENCE OF 1-87 FROM N\_A.  
 RC TISSUE-Placenta;  
 RX MEDLINE=94375010; Published=8088784;  
 RA Christiano A.M., Hoffman G.G., Chung-Honet L.C., Lee S., Cheng W., Ututto J.; Greenspan D.S.; "Structural organization of the human type VII collagen gene (COL7A1), composed of more exons than any previously characterized gene.";  
 RT Genomics 21:169-179(1994).  
 RL [8]  
 RN REVIEW ON DEB VARIANTS.  
 RP MEDLINE=94375848; Published=9375848;  
 RA Jaervikallio A., Pulkkinen L., Ututto J.; "Molecular basis of dystrophic epidermolysis bullosa: mutations in the type VII collagen gene (COL7A1).";  
 RT the type VII collagen gene (COL7A1).";  
 RL Hum. Mutat. 10:338-347(1997).  
 RN [9]  
 RP VARIANT RDEB LYS-2798.  
 RX MEDLINE=9329187; Published=8513325;  
 RA Christiano A.M., Greenspan D.S., Hoffman G.G., Zhang X., Tamai Y., Lin A.N., Dietz H.C., Hovnanian A., Ututto J.; "A missense mutation in type VII collagen in two affected siblings with recessive dystrophic epidermolysis bullosa.";  
 RT Nat. Genet. 4:62-66(1993).  
 RN [10]  
 RP VARIANT DDEB SER-2040.  
 RX MEDLINE=9422477; Published=8170945;  
 RA Christiano A.M., Rynneen M., Ututto J.; "Dominant dystrophic epidermolysis bullosa: identification of a gly-Ser substitution in the triple-helical domain of type VII collagen.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 91:3549-3553(1994).

- [11] RN PEB-DDEB CYS-2623.  
 RP VARIANT PEB-DDEB CYS-2623.  
 RX MEDLINE=96081220; PubMed=0541842;  
 RA Christiano A.M., Lee J.Y., Chen W.J., Laforgia S., Utto J.;  
 RT "Preterminal epidermolysis bullosa: genetic linkage to COL7A1 and  
 RT identification of a glycine-to-cysteine substitution in the triple-  
 helical domain of type VII collagen.";  
 RL Hum. Mol. Genet. 4:1579-1583(1995).  
 [12] RN DDEB ARG-2043.  
 RP VARIANT DDEB ARG-2043.  
 RX MEDLINE=95164985; PubMed=7861014;  
 RA Christiano A.M., Morricone A., Paradisi M., Angelo C., Mazzanti C.,  
 Cavalleri R., Utto J.;  
 RT "A glycine-to-arginine substitution in the triple-helical domain of  
 RT type VII collagen in a family with dominant dystrophic epidermolysis  
 RT bullosa.";  
 RL J. Invest. Dermatol. 104:438-440(1995).  
 [13] RN VARIANTS RDEB AND DDEB.  
 RP MEDLINE=96220218; PubMed=8644729;  
 RA Christiano A.M., McGrath J.A., Tan K.C., Utto J.;  
 RT "Glycine substitutions in the triple-helical region of type VII  
 collagen result in a spectrum of dystrophic epidermolysis bullosa  
 phenotypes and patterns of inheritance.";  
 An. J. Hum. Genet. 58:671-681(1996).  
 RN [14] RP VARIANT RDEB ARG-2575.  
 RX MEDLINE=96194068; PubMed=8592061;  
 RA Shimizu H., McGrath J.A., Christiano A.M., Nishikawa T., Utto J.;  
 RT "Molecular basis of recessive dystrophic epidermolysis bullosa:  
 RT genotype/phenotype correlation in a case of moderate clinical  
 severity.";  
 RL J. Invest. Dermatol. 106:119-124(1996).  
 RN [15] RP VARIANT RDEB ARG-1782.  
 RX MEDLINE=96183562; PubMed=8618018;  
 RA Christiano A.M., McGrath J.A., Utto J.;  
 RT "Influence of the second COL7A1 mutation in determining the  
 phenotypic severity of recessive dystrophic epidermolysis bullosa.";  
 J. Invest. Dermatol. 106:766-770(1996).  
 RN [16] RP VARIANT RDEB ASP-2073.  
 RX MEDLINE=96310789; PubMed=8757759;  
 RA Dunnill M.G.S., McGrath J.A., Richards A.J., Christiano A.M.,  
 RA Utto J., Rose P.M., Eady R.A.J.;  
 RT "Clinicopathological correlations of compound heterozygous COL7A1  
 mutations in recessive dystrophic epidermolysis bullosa.";  
 RL J. Invest. Dermatol. 107:171-177(1996).  
 RN [17] RP VARIANT RDEB W-1982; G-2008; A-2025; E-2049; G-2063 AND  
 RP 2575.  
 RX MEDLINE=97465605; PubMed=8326325;  
 RA Horvath A., Rochat A., Bodemer C., Petit E., Rivers C.A., Prost C.,  
 RA Fraitag S., Christiano A.M., Utto J., Lathrop M., Barrandon Y.,  
 RA de Prost Y.;  
 RT "Characterization of 18 new mutations in COL7A1 in recessive  
 RT dystrophic epidermolysis bullosa provides evidence for distinct  
 RT molecular mechanisms underlying defective anchoring fibril  
 RT formation.";  
 An. J. Hum. Genet. 61:599-610(1997).  
 RN [18] RP VARIANT RDEB ARG-1652.  
 RX MEDLINE=98106792; PubMed=9444387;  
 RA Cserhalmi-Friedman P.B., Karpatici S., Horvath A., Christiano A.M.;  
 RT "Identification of a glycine substitution and a splice site mutation  
 RT in the type VII collagen gene in a proband with mitis recessive  
 RT dystrophic epidermolysis bullosa.";  
 RL Arch. Dermatol. Res. 289:640-645(1997).  
 RN [19] RP VARIANTS DEB ARG-2009 AND ARG-2043.  
 RX MEDLINE=9738858; PubMed=9315684;  
 RA Winberg J.-O., Hamamci-Hausler N., Nilssen O., Anton-Lamprecht I.,  
 RA Naylor S.L., Kerbacher K., Zimmermann M., Krajci P.,  
 RA Gedde-Dahl T. Jr., Bruckner-Tuderman L.;  
 RT "Modulation of disease severity of dystrophic epidermolysis bullosa by  
 RT a splice site mutation in combination with a missense mutation in the  
 RT COL7A1 gene.";  
 RL Hum. Mol. Genet. 6:1125-1135(1997).  
 RN [20] RP VARIANTS DDEB ASP-1519; ASP-2006; GLU-2015 AND ARG-2034.  
 RX MEDLINE=98334662; PubMed=9668111;  
 RA Hammami-Hausler N., Schumann H., Raghunath M., Kilgus O., Luethi U.,  
 RA Lueger T., Bruckner-Tuderman L.;  
 RT "Some, but not all, glycine substitution mutations in COL7A1 result in  
 RT intracellular accumulation of collagen VII, loss of anchoring  
 RT fibrils, and skin blistering.";  
 RL J. Biol. Chem. 273:19228-19234(1998).  
 RN [21] RP VARIANTS DEB CYS-2008; ARG-2207 AND SER-2775.  
 RX MEDLINE=98410969; PubMed=9740253;  
 RA Kon A., Pulkkinen L., Ishida-Yamamoto A., Hashimoto I., Utto J.;  
 RT "Novel COL7A1 mutations in dystrophic forms of epidermolysis  
 RT bullosa.";  
 RL J. Invest. Dermatol. 111:534-537(1998).  
 RN [22] RP VARIANT RDEB ARG-1347.  
 RX MEDLINE=99019477; PubMed=9804332;  
 RA Terraccina M., Postoraro P., Schubert M., Sonnega G., Atzori F.,  
 RA Zambruno G., Bruckner-Tuderman L., Castiglione D.;  
 RT "Compound heterozygosity for recessive glycine substitution and a  
 RT splice site mutation in the COL7A1 gene causes an unusually mild form  
 RT of localized recessive dystrophic epidermolysis bullosa.";  
 RL J. Invest. Dermatol. 111:744-750(1998).  
 RN [23] RP VARIANT RDEB TRP-2034; VAL-2040; ARG-2043; ARG-2064 AND ASP-2713.  
 RX MEDLINE=99072663; PubMed=9856843;

```

Query Match          68.5%; Score 63; DB 1; Length 2944;
Matches   12; Conservative 0; MisMatches 0; Indels 3; Gaps 0;
QY      1 GVKGDKGNGPGWCGAP 15
Db     2386 GVKGDGLGLPGAP 2400

```

```

RESULT 8
CA34_HUMAN          STANDARD; PRM: 1670 AA.
ID CA34_HUMAN
AC 001955;
DT 01-OCT-1996 (Rel. 34, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Collagen alpha 3(IV) chain precursor (Goodpasture antigen).
GN COL4A3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OK NCBI_TAXID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
MEDLINE=94344949; PubMed=8083201;
RA Maruyama M., Leinonen A., Mochizuki T., Tryggvason K., Reeder S.T.;
RT "Complete primary structure of the human alpha 3(IV) collagen chain.  

  RT Coexpression of the alpha 3(IV) and alpha 4(IV) collagen chains in  

  RT human tissues.";  

  RL J. Biol. Chem. 269:23013-23017(1994).
RN [2]
RP REVISIONS
RA Leinonen A.;  

  RL Submitted (Oct-1998) to the EMBL/GenBank/DDBJ databases.
RN [3]
RN SEQUENCE OF 1386-1670 FROM N.A. AND PARTIAL SEQUENCE.
RP MEDLINE=93015826; PubMed=100291;
RX Quinones S., Bernal D., Garcia-Sogo M., Elena S.F., Saus J.;
```

RT \*Exon/intron structure of the human alpha 3(IV) gene encompassing the  
 RT Goodpasture antigen (alpha 3(IV)NC1). Identification of a potentially  
 RT antigenic region at the triple helix/NC1 domain junction. ";  
 RL J. Biol. Chem. 267:19780-19784(1992).  
 RN [4]  
 RP SEQUENCE OF 1453-1670 FROM N.A.  
 RX MEDLINE=91353570; PubMed=1883840;  
 RA "Sequence and localization of a partial cDNA encoding the human alpha  
 RC 3 chain of type IV collagen.";  
 RA Am. J. Hum. Genet. 49:545-554(1991).  
 RN [5]  
 RP SEQUENCE OF 1331-1570 FROM N.A.  
 RT TISSUE=Kidney;  
 RC MEDLINE=92147878; PubMed=1737849;  
 RA Turner N., Mason P.J., Brown R., Fox M., Povey S., Rees A.,  
 RA Pusey C.D.; "Molecular cloning of the human Goodpasture antigen demonstrates it  
 RT to be the alpha 3 chain of type IV collagen. ";  
 RL J. Clin. Invest. 89:592-601(1992).  
 RN [6]  
 RP SEQUENCE OF 1544-1670 FROM N.A.  
 RC TISSUE=Kidney;  
 RA Ding J.;  
 RA Submitted (JAN-1993) to the EMBL/GenBank/DBJ databases.  
 RL RN [7]  
 RP SEQUENCE OF 1439-1670, AND ALTERNATIVE SPlicing.  
 RC TISSUE=Kidney;  
 RX MEDLINE=9124597; PubMed=8294492;  
 RA Feng L., Xia Y., Wilson C.B.;  
 RT "Alternative splicing of the NC1 domain of the human alpha 3(IV)  
 collagen gene. Differential expression of mRNA transcripts that  
 predict three protein variants with distinct carboxyl regions. ";  
 RL J. Biol. Chem. 269:2342-2348(1994).  
 RN [8]  
 RP SEQUENCE OF 1-29 FROM N.A.  
 RX MEDLINE=98196854; PubMed=9537506;  
 RA Momota R., Sugimoto M., Oohashi T., Kigasawa K., Yoshioka H.,  
 RA Ninomiya Y.;  
 RT "Two genes, COL4A3 and COL4A4 coding for the human alpha3(IV) and  
 alpha4(IV) collagen chains are arranged head-to-head on chromosome  
 2936. ";  
 RL FEBS Lett. 424:11-16(1998).  
 RN [9]  
 RP ALTERNATIVE SPlicing.  
 RX MEDLINE=93280184; PubMed=8505332;  
 RA Bernal D., Quinones S., Saus J.;  
 RT "The human mRNA encoding the Goodpasture antigen is alternatively  
 RT spliced. ";  
 RL J. Biol. Chem. 268:12090-12094(1993).  
 RN [10]  
 RP VARIANT PRO-174.  
 RX MEDLINE=95078827; PubMed=7987301;  
 RA Lemmink H.H., Mochizuki T., van den Heuvel L.P.W.J., Schroeder C.H.,  
 RA Reenders S.T., Smeets L.J.M., van Oost B.A., Brunner H.G.,  
 RA Bartolets A., Monnens L.A.H., van der Putten H.,  
 RA "Mutations in the type IV collagen alpha 3 (COL4A3) gene in autosomal  
 RT recessive Alport syndrome. ";  
 RL Hum. Mol. Genet. 3:1269-1273(1994).  
 RN [11]  
 RP VARIANTS AS E-297; R-407; R-1167; E-1207; Q-1215; S-1277; T-  
 RP 1330; E-1334; E-1347; AND C-1661, AND VARIANTS R-43; E-162; Y-326; H-  
 RP 408; R-451; L-574; E-1269 AND P-1474.  
 RX MEDLINE=21064656; PubMed=11142255;  
 RA Heijnen L., Arrendel C., Forestier L., Cohen-Solal L., Mollet G.,  
 RA Gutierrez B., Stavrou C., Gubler M.C., Antignac C.;  
 RT "Structure of the human type IV collagen gene COL4A3 and mutations in  
 autosomal Alport syndrome. ";  
 RL J. Am. Soc. Nephrol. 12:97-106(2001).  
 CC -1- FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF  
 CC GLOMERULAR BASEMENT MEMBRANES (GBM), FORMING A 'CHICKEN-WIRE'  
 CC MESHWORK TOGETHER WITH LAMININS, PROTEOGLYCANS AND ENACTIN/  
 CC NIDOCIN.

CC -!- SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOFORMS, ALPHA 1(IV)-  
 CC ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE  
 CC WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.  
 CC -!- SUBCELLULAR LOCATION: CELL SURFACE (POTENTIAL).  
 CC -!- ALTERNATIVE PRODUCTS: AT LEAST 3 ISOFORMS; 1 (SHOWN HERE), 2/IV AND  
 CC 3/IV; ARE PRODUCED BY ALTERNATIVE SPlicing. THEY DIFFER IN THEIR  
 CC C-TERMINAL NC1 DOMAINS.  
 CC -!- TISSUE SPECIFICITY: ALPHA 3 AND ALPHA 4 TYPE IV COLLAGENS ARE  
 CC COLOCALIZED AND PRESENT ONLY IN BASEMENT MEMBRANES OF KIDNEY, EYE,  
 CC COCHLEA, LUNG AND BRAIN.  
 CC -!- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS  
 CC DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE  
 CC G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY  
 CC CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL  
 CC TRIPLE-HELICAL 7S DOMAIN.  
 CC -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPIDE REPEATING  
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
 CC -!- PTM: THE ALTERNATIVE SPliced FORM V CONTAINS AN ADDITIONAL  
 CC N-LINKED GLYCOSYLATION SITE.  
 CC -!- PROLINES AT THE THIRD POSITION OF THE TRIPEPIDE REPEATING  
 CC ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF  
 CC THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE  
 CC IV COLLAGENS.  
 CC -!- DISEASE: ANTIBODIES AGAINST THE NC1 DOMAIN OF ALPHA3(IV) MEDIATE  
 CC THE HUMAN AUTOIMMUNE DISEASE, GOODPASTURE SYNDROME, WHICH IS  
 CC CHARACTERIZED BY HEMATURIA AND PULMONARY HEMORRHAGE.  
 CC -!- DISEASE: DEFECTS IN COL4A3 ARE ASSOCIATED WITH THE TYPE I  
 CC AUTOSOMAL RECESSIVE FORM OF ALPORT SYNDROME, AN HEREDITARY  
 CC GLOMERULONPHRATHY CHARACTERIZED BY PROGRESSIVE RENAL FAILURE,  
 CC HEMATURIA AND DEAFNESS. THE RECESSIVE FORM OCCURS EQUALLY BETWEEN  
 CC MALES AND FEMALES.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC - the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.ebi-sib.ch/announce/>  
 CC or send an email to license@ebi-sib.ch).  
 CC -----  
 DR EMBL: X80031; CAA56335\_1;  
 DR EMBL: M9293; AAA21610\_1;  
 DR EMBL: S5579; AAC51556\_1;  
 DR EMBL: M81379; AAC51556\_1;  
 DR EMBL: L08650; AAC52044\_1;  
 DR EMBL: U02219; AAC18942\_1;  
 DR EMBL: U02520; AAC18943\_1;  
 DR EMBL: AB008495; BAA25064\_1;  
 DR MIM: 120070;  
 DR MIM: 203780;  
 DR MIM: 233450;  
 DR InterPro; IPR00087; Collagen.  
 DR Pfam; PF01411; C4\_2;  
 DR Pfam; PF01391; Collagen\_21.  
 DR Pfam; PF003923;  
 DR SMART; SM00111; C4\_2;  
 DR Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
 KW Glycoprotein; Basement membrane; Collagen; Signal; Cell adhesion;  
 KW Alternative splicing; Polymorphism; Disease mutation;  
 KW Airport syndrome.  
 FT SIGNAL 1 28 POTENTIAL.  
 FT CHAIN 29 1670 COLLAGEN ALPHA 3 (IV) CHAIN.  
 FT DOMAIN 29 42 7S DOMAIN.  
 FT DOMAIN 43 1438 TRIPLE-HELICAL REGION.  
 FT DOMAIN 1439 1670 NONHELICAL REGION (NC1) (GOODPASTURE  
 FT ANTIGEN) (BY SIMILARITY).  
 FT EPIPOLE (RECOGNIZED BY GOODPASTURE  
 FT ANTIBODIES).  
 FT CLEAVAGE (BY COLLAGENASE)  
 FT (BY SIMILARITY).  
 SITE 791 793 CELL ATTACHMENT SITE (POTENTIAL).  
 SITE 996 998 CELL ATTACHMENT SITE (POTENTIAL).

FT SITE 1154 1156 CELL ATTACHMENT SITE (POTENTIAL).  
 FT SITE 1306 1308 CELL ATTACHMENT SITE (POTENTIAL).  
 FT SITE 1345 1347 CELL ATTACHMENT SITE (POTENTIAL).  
 FT SITE 1432 1434 CELL ATTACHMENT SITE (POTENTIAL).  
 FT CARBOHYD 253 253 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT MOD\_RES 1435 1435 PHOSPHORYLATION (BY SIMILARITY).  
 FT MOD\_RES 1437 1437 PHOSPHORYLATION (BY SIMILARITY).  
 FT DISULFID 1460 1551 OR 1548 (BY SIMILARITY).  
 FT DISULFID 1493 1548 OR 1551 (BY SIMILARITY).  
 FT DISULFID 1505 1511 BY SIMILARITY.  
 FT DISULFID 1570 1665 OR 1662 (BY SIMILARITY).  
 FT DISULFID 1604 1662 OR 1665 (BY SIMILARITY).  
 FT DISULFID 1616 1622 BY SIMILARITY.  
 FT VARSPLIC 1586 1670 VARSPLIC  
 FT VARSPLIC 1488 1670 VARSPLIC  
 FT VARSPLIC 1488 1670 /FTid=VAR\_011202.  
 FT VARIANT 43 43 G -> R.  
 FT VARIANT 162 162 G -> E.  
 Query Match 67.4%; Score 62; DB 1; Length 1670;  
 Best Local Similarity 66.7%; Pred. No. 0.16; Mismatches 2;  
 Matches 10; Conservative 2; Pred. No. 0.16; Indels 3; Pred. No. 0.046;  
 Qy 1 GVKDCKGNPGWPSAP 15 Mismatches 1; Gaps 0; Gaps 0;  
 DE Minor capsid protein (Protein P5).  
 GN 1189 GAKGDRGAPGFFGFLP 1203  
 Db 123 GIKGDKGDPGAPG 135  
 RESULT 9  
 COAS\_BPPRD STANDARD; PRT; 339 AA.  
 AC P22536; RN [1]  
 DT 01-AUG-1991 (Rel. 19, Created)  
 DT 01-DEC-1992 (Rel. 24, Last sequence update)  
 DE Minor capsid protein (Protein P5).  
 V.  
 OS Bacteriophage PRD1.  
 OS Viruses; dsDNA viruses, no RNA stage; Tectiviridae; Tectivirus.  
 OX NCBI\_TAXID=1058;  
 RN [1] SEQUENCE FROM N.A., AND SEQUENCE OF 1-5.  
 RP MEDLINE-90320115; PubMed=2156741;  
 RA Bamford J.K.H., Bamford D.H.;  
 RT "Capsomer proteins of bacteriophage PRD1, a bacterial virus with a  
 membrane";  
 RT virology 177:445-451(1990).  
 RL [2] FUNCTION: THE SHORT COLLAGEN-LIKE REGION IS PROPOSED TO ACT AS A  
 RP COLLAGENOUS REGION.  
 RP MEDLINE-90206069; PubMed=2320123;  
 RA Bamford D.H., Bamford J.K.H.;  
 RT "Collagenous Proteins multiply.";  
 RL Nature 344:497-497(1990).  
 CC -1- FUNCTION: THE SHORT COLLAGEN-LIKE REGION IS PROPOSED TO ACT AS A  
 CC TRIMERIZATION SIGNAL LEADING TO THE FULFILLMENT OF THE SYMMETRY  
 CC REQUIREMENT OF THE MINOR CAPSOMER.  
 CC -1- SUBUNIT: THE PRD1 VIRION IS COMPOSED OF AN ICOSAHEDRAL PROTEIN  
 CC COAT, AN INNER PROTEIN-LIPID MEMBRANE, AND A DSDNA GENOME WHICH  
 CC IS LOCATED INSIDE THE LIPID VESICLE. THE PROTEIN COAT CONSISTS  
 CC MAINLY OF PROTEIN P3. THE MINOR PROTEIN P5 IS ALSO CONSIDERED TO  
 CC CONSTITUTE PART OF THE PROTEIN COAT.  
 CC -1- SIMILARITY: CONTAINS A COLLAGENOUS DOMAIN.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>).  
 CC or send an email to license@isb-sib.ch).  
 CC  
 DR EMBI; MG9077; AAA32460 1; --.  
 DR EMBI; M55568; AAA32446 1; --.  
 DR B46345; B46345.  
 PIR; Coat protein.  
 KW  
 FT INIT\_MET 0 0  
 FT DOMAIN 1 121 DOMAIN-1  
 FT DOMAIN 122 140 COLLAGEN-LIKE.  
 FT DOMAIN 141 339 DOMAIN-2.  
 SQ 339 AA: 34318 MW: 184BCE77D08944F3 CRC64;  
 Query Match 66.3%; Score 61; DB 1; Length 339;  
 Best Local Similarity 75.9%; Pred. No. 0.046;  
 Matches 10; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 Qy 1 GVKDCKGNPGWPG 13  
 Db 123 GIKGDKGDPGAPG 135  
 RESULT 10  
 CA1F\_HUMAN STANDARD; PRT; 1603 AA.  
 ID CA1F\_HUMAN STANDARD; PRT; 1603 AA.  
 AC 007092; RN [1]  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DE 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Collagen alpha 1(XVI) chain precursor.  
 GN  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TAXID=9606;  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-92333339; PubMed=1631157;  
 RA Pan T.-C., Zheng R. Z., Mattei M.-G., Timpl R., Chu M.-L.;  
 RT "Cloning and chromosomal location of human alpha 1(XVI) collagen."  
 RL PROC. NATL. ACAD. SCI. U.S.A. 89:6565-6569(1992).  
 RN [2]  
 RP SEQUENCE OF 418-1603 FROM N.A.  
 RC TISSUE=placenta;  
 RX MEDLINE-93201361; PubMed=1284248;  
 RA Yamaguchi N., Kimura S., McBride O.W., Hori H., Yamada Y.,  
 RA Kanamori T., Yamakoshi H., Nagai Y.;  
 RT "Molecular cloning and partial characterization of a novel collagen  
 chain, alpha 1(XVI), consisting of repetitive collagenous domains and  
 RT cysteine-containing non-collagenous segments.";  
 RL J. Biochem. 112:856-863(1992).  
 CC  
 CC -1- FUNCTION: THE NUMEROUS INTERRUPTIONS IN THE TRIPLE HELIX MAY MAKE  
 CC THIS MOLECULE EITHER ELASTIC OR FLEXIBLE.  
 CC -1- TISSUE SPECIFICITY: IN THE PLACENTA, WHERE IT IS FOUND IN THE  
 CC AMNIOTIC, A MEMBRANOUS TISSUE LINING THE AMNIOTIC CAVITY. WITHIN THE  
 CC AMNIOT, IT IS FOUND IN AN ACCELLULAR, RELATIVELY DENSE LAYER OF A  
 CC COMPLEX NETWORK OF RETICULAR FIBERS, ALSO LOCATED TO A FIBROBLAST  
 CC LAYER BEHIND THIS DENSE LAYER. EXISTS IN TISSUES IN ASSOCIATION  
 CC WITH OTHER TYPES OF COLLAGEN.  
 CC -1- DEVELOPMENTAL STAGE: TRANSIENTLY ELEVATED EXPRESSION DURING  
 CC GESTATION, AND DECREASE AT TERM.  
 CC -1- DOMAIN: THIS SEQUENCE DEFINES EIGHTEEN DIFFERENT DOMAINS, NINE  
 CC TRIPLE-HELICAL DOMAINS (COL9 TO COL1) AND TEN NONTRIPLE-HELICAL  
 CC DOMAINS (NC10 TO NC1).  
 CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRipeptIDE REPEATING  
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
 CC -1- SIMILARITY: BELONGS TO THE FIBRIL-ASSOCIATED COLLAGENS WITH  
 CC INTERRUPTED HELICES (FACIT) FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -	
CC	the European Bioinformatics Institute. There are no restrictions on its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> )	
CC	or send an email to license@isb-sib.ch).	
DR	EMBL; M92642; AAA58427.1; -.	
DR	EMBL; S57132; AAB25797.1; -.	
DR	PIR; S23810; S23810.	
DR	MIM: 120326; -.	
DR	InterPro: IPR00087; collagen.	
DR	InterPro: IPR03129; TSPN.	
DR	Pfam; PF01391; Collagen; 16.	
DR	SMART; SM00210; TSPN; 1.	
FT	Extracellular matrix; Connective tissue; Collagen; Hydroxylation;	
FT	Repeat; Signal; 21	
FT	SIGNAL	
FT	CHAIN 22 1603	
FT	DOMAIN 374 374	
FT	DOMAIN 375 505	
FT	DOMAIN 506 520	
FT	DOMAIN 521 554	
FT	DOMAIN 555 571	
FT	DOMAIN 572 630	
FT	DOMAIN 631 651	
FT	DOMAIN 652 722	
FT	DOMAIN 723 737	
FT	DOMAIN 738 875	
FT	DOMAIN 876 886	
FT	DOMAIN 887 938	
FT	DOMAIN 939 972	
FT	DOMAIN 973 987	
FT	DOMAIN 988 1010	
FT	DOMAIN 1011 1432	
FT	DOMAIN 1433 1471	
FT	DOMAIN 1472 1577	
FT	DOMAIN 1578 1603	
FT	CONFLICT 418 420	
FT	CONFLICT 537 537	
FT	CONFLICT 1160 1160	
FT	CONFLICT 1163 1163	
FT	CONFLICT 1165 1165	
SQ	SEQUENCE 1603 AA; 157692 MW; E27D9A1DAE59BA37 CRC64;	
Query Match	66.3%; Score 61; DB 1; Length 1603;	
Best Local Similarity	73.3%; Pred. No. 0.21; Mismatches 11; Conservative 0; Indels 4; Gaps 0;	
Matches	1; Last sequence update)	
OY	1 GVKGDKGNPGWGPAGP 15	
Db	681 GQKGDAGNPGDPGTP 695	
RESULT 11		
CA44_HUMAN	STANDARD; PRT; 1690 AA.	
ID	CA44_HUMAN	
AC	P53420; 01-OCT-1996 (Rel. 34, Created)	
DT	01-OCT-1995 (Rel. 34, Last sequence update)	
DT	01-MAR-2002 (Rel. 41, Last annotation update)	
DE	Collagen alpha 4(IV) chain precursor.	
OS	Homo sapiens (Human).	
RX	MEDLINE=99011253; PubMed=8792860;	
RA	Boye E., Mollet G., Forestier L., Cohen-Solal L., Heidet L., Lemmink H.H., Nillesius W.N., Mochizuki T., Schroeder C.H., Brunner H.-P., van Oost B.A., Monnens L.A.H., Smeets H.J.M., Benign familial hematuria due to mutation of the type IV collagen alpha4 gene.;	
RT	J Clin. Invest. 98:1114-1118(1996).	
RL	Am. J. Hum. Genet. 63:1320-1340(1998).	
CC	-I- FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF GLOMERULAR BASEMENT MEMBRANES (GBM), FORMING A 'CHICKEN-WIRE' MESHWORK TOGETHER WITH LAMININS, PROTEOGLYCANS AND ENACTIN/ NITROGEN. SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOFORMS, ALPHA 1 (IV)-	
CC	-I- HOMO sapiens (Human).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
OC	OX Mammalia; Taxid:9606;	
RA	[1]	
RP	SEQUENCE FROM N.A.	
RC	TISSUE-Kidney;	
RX	MEDLINE=9501445; PubMed=7523402;	
RA	Leinonen A., Maruyama M., Mochizuki T., Tryggvason K., Reeder S.T.; "Complete primary structure of the human type IV collagen alpha 4(IV) chain. Comparison with structure and expression of the other alpha	
RN	(IV) chains"; J. Biol. Chem. 269:26172-26177(1994).	
RL	RN [2]	
RP	SEQUENCE OF 1-23 FROM N.A.	
RX	MEDLINE=98196854; PubMed=9357506;	
RA	Momota R., Sugimoto M., Oohashi T., Kigasawa K., Yoshioka H., Niromiya Y.;	
RT	"Two genes, COLA3 and COLA4 coding for the human alpha3(IV) and alpha4(IV) collagen chains are arranged head-to-head on chromosome 2q36.", FEBS Lett. 424:11-16(1998).	
RL	RN [3]	
RP	SEQUENCE OF 1219-1690 FROM N.A.	
RC	TISSUE-Eye;	
RX	MEDLINE=9337407; PubMed=8365481;	
RA	Sugimoto M., Ooashi T., Yoshikawa H., Masuo N., Niromiya Y.; "cDNA isolation and partial gene structure of the human alpha 3(IV) collagen chain.", FEBS Lett. 330:122-128(1993).	
RN	RN [4]	
RP	SEQUENCE OF 1407-1507 FROM N.A.	
RX	MEDLINE=93054733; PubMed=142974;	
RA	Kamagata Y., Mattei M.-G., Niromiya Y., "Isolation and sequencing of cDNAs and genomic DNAs encoding the alpha 4 chain of basement membrane collagen type IV and assignment of the gene to the distal long arm of human chromosome 2."; J. Biol. Chem. 267:23753-23758(1992).	
RL	RN [5]	
RP	REVIEW ON VARIANTS.	
RX	MEDLINE=97338662; PubMed=919522;	
RA	Lemmink H.H., Schroeder C.H., Monnens L.A.H., Smeets H.J.M.; "The clinical spectrum of type IV collagen mutations.," Hum. Mutat. 9:477-490(1997).	
RL	RN [6]	
RP	VARIANT AS SER-1201.	
RX	MEDLINE=95078927; PubMed=7987396;	
RA	Mochizuki T., Lemmink H.H., Maruyama M., Antignac C., Gubler M.-C., Pirson Y., Verelle-Dumoulin C., Chan B., Schroeder C.H., Smeets H.J.M., Reeder S.T.; "Identification of mutations in the alpha 3(IV) and alpha 4(IV) collagen genes in autosomal recessive Alport syndrome.," RT Nat. Genet. 8:77-82(1994).	
RL	RN [7]	
RP	VARIANT BRH GLU-897.	
RX	MEDLINE=96379660; PubMed=8787673;	
RA	Lemmink H.H., Nillesius W.N., Mochizuki T., Schroeder C.H., Brunner H.-P., van Oost B.A., Monnens L.A.H., Smeets H.J.M., Benign familial hematuria due to mutation of the type IV collagen alpha4 gene.;	
RT	J Clin. Invest. 98:1114-1118(1996).	
RL	RN [8]	
RP	VARIANT AS AND VARIANTS.	
RX	MEDLINE=99011253; PubMed=8792860;	
RA	Boye E., Mollet G., Forestier L., Cohen-Solal L., Heidet L., Lemmink H.H., Nillesius W.N., Mochizuki T., Schroeder C.H., Brunner H.-P., van Oost B.A., Monnens L.A.H., Smeets H.J.M., Benign familial hematuria due to mutation of the type IV collagen alpha4 gene.;	
RT	"The clinical spectrum of type IV collagen mutations.," Hum. Mutat. 9:477-490(1997).	
RL	Am. J. Hum. Genet. 63:1320-1340(1998).	
CC	-I- FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF GLOMERULAR BASEMENT MEMBRANES (GBM), FORMING A 'CHICKEN-WIRE' MESHWORK TOGETHER WITH LAMININS, PROTEOGLYCANS AND ENACTIN/ NITROGEN. SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOFORMS, ALPHA 1 (IV)-	
CC	-I- HOMO sapiens (Human).	



CC IV COLLAGENS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration -

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (see <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC EMBL; M67507; AAAI18014.1; -.

CC InterPro; IPR01442; C4.

CC PIR; S16366; S16366.

CC InterPro; IPR00087; Collagen.

CC Pfam; PF01391; Collagen; 23.

CC Prodrom; PDO03923; C4; 2.

CC SMART; SM00111; C4; 2.

CC DR SM00111; C4; 2.

CC DR EMBL; M67507; AAAI18014.1; -.

CC DR InterPro; IPR01442; C4.

CC DR PIR; S16366; S16366.

CC DR InterPro; IPR00087; Collagen.

CC Pfam; PF01391; Collagen; 23.

CC Prodrom; PDO03923; C4; 2.

CC KW Hydroxylation; Connective tissue; Basement membrane; Repeat; Collagen;

CC Alternative splicing; Glycoprotein; Signal.

FT SIGNAL 1; 26

FT CHAIN 27 1763

FT DOMAIN 27 42

FT DOMAIN 43 1529

FT DOMAIN 1530 1763

FT DISULFID 1548 1637

FT DISULFID 1581 1634

FT DISULFID 1593 1599

FT DISULFID 1656 1752

FT DISULFID 1690 1749

FT DISULFID 1702 1709

FT CARBOHYD 126 126

FT CARBOHYD 249 249

FT VARSPLIC 230 266

FT ISOFORM II 1763 AA; 168526 MW; 304F528BC06AAE0D CRC64;

CC Query Match 66.3%; Score 61; DB 1; Length 1763;

CC Best Local Similarity 73.3%; Pred. No. 0; 23; Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

CC QY 1 GVKGDKGNPGWGP 15

CC Db 760 GLPGMKGNPGLGAP 774

RESULT 13

PSPD\_BOVIN STANDARD; PRT; 369 AA.

ID P35246; 01-FEB-1994 (Rel. 28, Created)

DT 15-JUL-1999 (Rel. 38, Last sequence update)

DE pulmonary surfactant-associated protein D precursor (SP-D) (PSP-D).

OS SFTPDB OR SFTP4.

OS Bos taurus (Bovine).

OC Bovidae; Bovinae; Bos.

OC Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi; Bivalvia; Bovinae; Bos.

OX NCBI\_TaxID=9913; [1]

RP SEQUENCE FROM N.A., AND SEQUENCE OF 208-247.

RC TISSUE=Lung;

RX MEDLINE=93170855; PubMed=043642;

RX Lim B.L., Lu J.J., Reid K.B.M.; RT "structural similarity between bovine conglutinin and bovine lung surfactant protein D and demonstration of liver as a site of synthesis of conglutinin";

RL Immunology 78:159-165(1993). -1- FUNCTION: CONTRIBS TO THE LUNG'S DEFENSE AGAINST INHALED MICROORGANISMS. BINDS STRONGLY MALTOSE RESIDUES AND TO A LESSER

CC EXTENT OTHER ALPHA-GLUCOSYL MOIETIES. IT COULD PARTICIPATE IN THE EXTRACELLULAR REORGANIZATION OR TURNOVER OF PULMONARY SURFACTANT.

CC -!- SUBUNIT: OLIGOMERIC COMPLEX OF 4 SET OF HOMOTRIMERS.

CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR.

CC -!- MISCELLANEOUS: PULMONARY SURFACTANT CONSISTS OF 90% LIPID AND 10% PROTEIN. THERE ARE 4 SURFACTANT ASSOCIATED PROTEIN: 2 COLLAGENOUS, CARBOHYDRATE BINDING GLYCOPROTEINS (SP-A AND SP-D) AND 2 SMALL HYDROPHOBIC PROTEINS (SP-B AND SP-C).

CC -!- SIMILARITY: CONTAINS A COLLAGENOUS DOMAIN.

CC -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.

CC SIMILARITY: CONTAINS 1 C-TYPE LECTIN.

CC DR EMBL; X75911; CAAS3510.1; -.

CC DR PIR; S33603; S33603.

CC DR HSSP; P35247; 1B08.

CC DR InterPro; IPR00087; Collagen.

CC DR InterPro; IPR01304; Lectin\_c.

CC DR Pfam; PF01391; Collagen; 3.

CC DR Pfam; PF00059; Lectin\_c; 1.

CC DR SMART; SM0014; CLECT; 1.

CC DR PROSITE; PS00615; C\_TYPELECTIN\_1; 1.

CC DR PROSITE; PS50041; C\_TYPELECTIN\_2; 1.

CC KW Glycoprotein; Calcium; Surface film; Gaseous exchange; Hydroxylation; Signal; Lectin; Collagen; Repeat; Coiled coil.

CC FT SIGNAL 1; 20

CC FT CHAIN 21 369

CC FT DOMAIN 46 216

CC FT DOMAIN 217 248

CC FT DOMAIN 273 369

CC FT DISULFID 275 367

CC FT DISULFID 345 359

CC FT CARBOHYD 90 90

CC FT MOD\_RES 78 78

CC FT MOD\_RES 87 87

CC FT MOD\_RES 95 96

CC FT MOD\_RES 99 99

CC FT MOD\_RES 165 165

CC FT MOD\_RES 171 171

CC SQ SEQUENCE 369 AA; 37361 MW; 07D88B24E0AEB2E3 CRC64;

CC Query Match 65.2%; Score 60; DB 1; Length 369;

CC Best Local Similarity 66.7%; Pred. No. 0; 07; Matches 10; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

CC QY 1 GVKGDKGNPGWGP 15

CC Db 157 GLKGERGAGEPGAP 171

RESULT 14

CA44\_RABBIT STANDARD; PRT; 623 AA.

ID CA44\_RABBIT

AC P55787; 01-NOV-1997 (Rel. 35, Last sequence update)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 30-MAY-2000 (Rel. 39, Last annotation update)

DE Collagen alpha 4(IV) chain (Fragment).

GN COL4A4.

OS Oryctolagus cuniculus (Rabbit).

OC Eutheria; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.

OX NCBI\_TaxID=9986; [1]

RP SEQUENCE FROM N.A. TISSUE=Corneal endothelium;

RC

RA	MEDLINE-93054733;	PubMed=1429714;	Rattus norvegicus (Rat).
Kanegata Y., Mattel M.-G., Ninomiya Y.	"Isolation and sequencing of cDNAs and genomic DNAs encoding the alpha 4 chain of basement membrane collagen type IV and assignment of the gene to the distal long arm of human chromosome 2.;"		
RT	J. Biol. Chem. 267:23753-23758(1992).		
RT	-1- FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF GLOMERULAR BASEMENT MEMBRANES (GBM), FORMING A "CHICKEN-WIRE" MESHWORK TOGETHER WITH LAMININS, PROTEOGLYCANS AND ENTHACTIN/NIDOCIN.		
RT	-1- SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOFORMS, ALPHA 1(IV)-ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE WITH 2 OTHER CHAINS TO GENERATE TYPE IV COLLAGEN NETWORK.		
RT	-1- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS DOMAIN (NC1) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE G-X-Y REPORTS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL TRIPLE-HELICAL 7S DOMAIN.		
RT	-1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPETIDE REPEATING UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.		
RT	-1- PTM: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF THESE, LOCATED IN THE NC1 DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE IV COLLAGENS.		
RT	-1- SIMILARITY: TO OTHER TYPE IV COLLAGENS.		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to license@isb-sib.ch).		
CC	EMBL; L01477; ; NOT_ANNOTATED_CDS.		
CC	InterPro; IPR00087; Collagen.		
DR	Pfam; PF01391; Collagen; 5.		
DR	ProDom; PD00323; C4; 2.		
DR	SMART; SM00111; C4; 2.		
KW	Extracellular matrix; Connective tissue; Repeat; Hydroxylation; Glycoprotein; Basement membrane; Collagen; Cell adhesion.		
FT	NON_TER 1 <1 392 TRIPLE-HELICAL REGION.		
FT	DOMAIN 393 623 NONHICAL REGION (NC1).		
FT	DISULFID 413 502 OR 499 (BY SIMILARITY).		
FT	DISULFID 446 499 OR 502 (BY SIMILARITY).		
FT	DISULFID 458 464 BY SIMILARITY.		
FT	DISULFID 521 619 OR 619 (BY SIMILARITY).		
FT	DISULFID 555 616 OR 619 (BY SIMILARITY).		
FT	DISULFID 567 574 BY SIMILARITY.		
SQ	SEQUENCE 623 AA; 62393 MW; CCBG9BB31242EE82 CRC64;		
Query Match	65.2%; Score 60; DB 1; Length 623;		
Best Local Similarity	76.9%; Pred. No. 0.12; 10; Conservative		
Matches	1; Mismatches 2; Indels 0; Gaps 0;		
Oy	1 GVKGDKGNPGWPG 13		
Db	139 GIKGDKGEPGSPG 151		
RESULT 15			
CA13-RAT			
ID			
AC	P13941; 070604;		
DT	01-JAN-1990 (Rel. 13, Created)		
DT	01-JUN-1994 (Rel. 29, Last sequence update)		
DT	16-OCT-2001 (Rel. 40, Last annotation update)		
DT	Collagen alpha 1(III) chain (Fragment).		
COLLAB.	COLLAB.		
Query Match	65.2%; Score 60; DB 1; Length 636;		
Best Local Similarity	71.4%; Pred. No. 0.12; 10; Conservative		
Matches	3; Mismatches 1; Indels 0; Gaps 0;		
Oy	1 GVKGDKGNPGWPG 14		

Db: 30 GVKGGERGSPGGPA 43

RESULT 16  
 CAL3\_BOVIN ID CAL3\_BOVIN STANDARD; PRT; 1049 AA.  
 AC P04238;  
 DT 20-MAR-1987 (Rel. 04, Created)  
 DT 20-MAR-1987 (Rel. 04, Last sequence update)  
 DE 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Collagen alpha 1(III) chain.  
 GN COL3A1  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;  
 Bovidae; Bovinae; Bos.  
 NCBI\_TAXID=9913;  
 [1]  
 RP SEQUENCE OF 1-242; PubMed=488906;  
 RX MEDLINE=80026026;  
 RA Fietzek P.P., Alimann H., Rauterberg J., Henkel W., Wachter E.,  
 RA Kuhn K.;  
 RT "The covalent structure of calf skin type III collagen. I. The amino acid sequence of the amino terminal region of the alpha 1(III) chain (positions 1-222).";  
 RT Hoppe-Seyler's Z. Physiol. Chem. 360:809-820(1979).  
 RL [2]  
 RP SEQUENCE OF 243-422;  
 RX MEDLINE=80026027; PubMed=488907;  
 RA Dewes H., Fietzek P.P., Kuhn K.;  
 RT "The covalent structure of calf skin type III collagen. II. The amino acid sequence of the cyanogen bromide peptide alpha 1(III)CB1,8,10,2 (positions 223-402).";  
 RT Hoppe-Seyler's Z. Physiol. Chem. 360:821-832(1979).  
 RL [3]  
 RP SEQUENCE OF 423-571.  
 RX MEDLINE=80026028; PubMed=488908;  
 RA Bentz H., Fietzek P.P., Kuhn K.;  
 RT "The covalent structure of calf skin type III collagen. III. The amino acid sequence of the cyanogen bromide peptide alpha 1(III)CB4 (positions 403-551).";  
 RT Hoppe-Seyler's Z. Physiol. Chem. 360:833-840(1979).  
 RL [4]  
 RP SEQUENCE OF 572-808;  
 RX MEDLINE=80026029; PubMed=488909;  
 RA Lang H., Gianville R.W., Fietzek P.P., Kuhn K.;  
 RT "The covalent structure of calf skin type III collagen. IV. The amino acid sequence of the cyanogen bromide peptide alpha 1(III)CB5 (positions 552-788).";  
 RT Hoppe-Seyler's Z. Physiol. Chem. 360:841-850(1979).  
 RL [5]  
 RP SEQUENCE OF 809-947.  
 MEDLINE=80026030; PubMed=488910;  
 RA Dewes H., Fietzek P.P., Kuhn K.;  
 RT "The covalent structure of calf skin type III collagen. V. The amino acid sequence of the cyanogen bromide peptide alpha 1(III)CB9A (position 789-927).";  
 RT Hoppe-Seyler's Z. Physiol. Chem. 360:851-860(1979).  
 RL [6]  
 RP SEQUENCE OF 948-1049.  
 RX MEDLINE=80026031; PubMed=488911;  
 RA Allmann H., Fietzek P.P., Gianville R.W., Kuhn K.;  
 RT "The covalent structure of calf skin type III collagen. VI. The amino acid sequence of the carboxy-terminal cyanogen bromide peptide alpha 1(III)CB9B (positions 928-1028).";  
 RT Hoppe-Seyler's Z. Physiol. Chem. 360:861-868(1979).  
 -I- FUNCTION: COLLAGEN TYPE III OCCURS IN MOST SOFT CONNECTIVE TISSUES  
 ALONG WITH TYPE I COLLAGEN.  
 CC -I- SUBUNIT: TRIMERS OF IDENTICAL ALPHA 1(III) CHAINS. THE CHAINS ARE LINKED TO EACH OTHER BY INTERCHAIN DISULFIDE BONDS. TRIMERS ARE ALSO CROSS-LINKED VIA HYDROXYLINES.  
 CC -I- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPÉPTIDE REPEATING UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.

DR PIR: A02862; CCBO75;  
 DR InterPro: IPR000087; Collagen.  
 DR Pfam: PF01391; Collagen; 17.  
 DR PROSITE: PS01208; VWFC; PARTIAL.  
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
 GLYCOPROTEIN; Collagen.  
 FT DOMAIN 1 14 NONHELCICAL REGION (N-TERMINAL).  
 FT DOMAIN 15 1040 TRIPLE-HELICAL REGION.  
 FT DOMAIN 1041 1049 NONHELCICAL REGION (C-TERMINAL).  
 MOD\_RES 95 95 HYDROXYLATION.  
 FT MOD\_RES 107 107 HYDROXYLATION.  
 FT CARBOHYD 950 950 O-LINKED (GAL. . .).  
 FT CARBOHYD 950 950 O-LINKED (GAL. . .).  
 FT DISUFRID 1040 1040 INTERCHAIN.  
 FT DISUFRID 1041 1041 INTERCHAIN.  
 SQ SEQUENCE 1049 AA; 93651 MW; 8EBC33DIC66ECC9A3 CRC64;

Query Match 65.2%; Score 60; DB 1; length 1049;  
 Best Local Similarity 71.4%; Pred. No. 0.2;  
 Matches 10; Conservative; 3; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 GVKGGERGSPGGPA 14  
 ||||:||||| 702 GVKGGERGSPGGPA 715

Db RESULT 17  
 CAL3\_CHICK ID CAL3\_CHICK STANDARD; PRT; 1262 AA.  
 AC P12105; P79758; P79759; 090794; 092029;  
 DT 01-OCT-1989 (Rel. 12, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Collagen alpha 1(III) chain precursor (Fragments).  
 GN COL3A1  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
 OC Gallus.  
 OC NCBI\_TAXID=9031;

RN [1]  
 RP SEQUENCE OF 1-886 FROM N.A.  
 RC TISSUE=Kidney;  
 RX MEDLINE=94266842; PubMed=8206952;  
 RA Nah H.-D., Niu Z., Adams S.L.;  
 RT "An alternative transcript of the chick type III collagen gene that does not encode type III collagen.";  
 RL J. Biol. Chem. 269:16443-16448(1994).  
 RN [2]  
 RP SEQUENCE OF 29-96; 332-397; 431-484; 503-535 AND 869-976 FROM N.A.  
 RX MEDLINE=84270596; PubMed=6547770;  
 RA Yamada Y., Liao G., Mudryj M., Obici S., de Crombrugge B.;  
 RT "Conservation of the sizes for one but not another class of exons in two chick collagen genes.";  
 RL Nature 310:333-337(1984).  
 RN [3]  
 RP SEQUENCE OF 977-1262 FROM N.A.  
 RX MEDLINE=83220816; PubMed=6856474;  
 RA Yamada Y., Kuhn K., de Crombrugge B.;  
 RT "A conserved nucleotide sequence, coding for a segment of the C-propeptide, is found at the same location in different collagen genes.";  
 RT Nucleic Acids Res. 11:2733-2744(1983).  
 CC -I- FUNCTION: COLLAGEN TYPE III OCCURS IN MOST SOFT CONNECTIVE TISSUES  
 CC -I- SUBUNIT: TRIMERS OF IDENTICAL ALPHA 1(III) CHAINS. THE CHAINS ARE LINKED TO EACH OTHER BY INTERCHAIN DISULFIDE BONDS. TRIMERS ARE ALSO CROSS-LINKED VIA HYDROXYLINES.  
 CC -I- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPÉPTIDE REPEATING UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.

CC	-1 - SIMILARITY: CONTAINS 1 WFEC DOMAIN.	DT	16-OCT-2001 (Rel. 40, last annotation update)
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration -	DE	Collagen alpha 1(III) chain precursor.
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -	GN	COL3A1.
CC	use by non profit institutions as long as its content is in no way	OS	Homo sapiens (Human).
CC	modified and this statement is not removed. Usage by and for commercial/	OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC	entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> ) or send an email to license@isb-sib.ch).	OC	Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo;
CC		OX	NEBIL TAXID=9606;
CC		RN	[1]
EMBL: U07973; AAH83407.1; -	SPQUENCE FROM N.A.	RP	SPQUENCE FROM N.A.
DR	EMBL; X00822; CAB52686.1; -	RC	TISSUE-Skin fibroblast;
DR	EMBL; X00823; CAB52686.1; JOINED.	RX	Medline=89310838; Pubmed=2764886;
DR	EMBL; X00826; CAA25397.1; ALT_Seq.	RA	Ala-Kotko L., Kontusaari S., Baldwin C.T., Kulvantieni H.,
DR	EMBL; X00825; CAA25397.1; JOINED.	RA	Prockop D.J.;
DR	EMBL; X00827; CAA25398.1; -	RT	Structure of cDNA clones coding for the entire prepro alpha 1 (III) chain of human type III procollagen. Differences in protein structure from type I procollagen and conservation of codon preferences.;
DR	EMBL; X00828; CAA25399.1; -	RT	Janecko R.A., Ramirez F.;
DR	EMBL; X00830; CAA25402.1; -	RT	"Nucleotide and amino acid sequences of the entire human alpha 1 (III) collagen.";
DR	EMBL; X00831; CAA25402.1; -	RL	Biochem. J. 260:509-516(1989).
DR	EMBL; X00832; AAD15299.1; -	RN	[2]
DR	EMBL; M36662; AAA18519.1; ALT_SEQ.	RP	SEQUENCE OF 149-1225 FROM N.A.
DR	PIR: A05269.	RX	Medline=89386015; Pubmed=2780304;
DR	InterPro; IPK00087; Collagen.	RA	Seyer J.M., Kang A.H.;
DR	InterPro; IPR00085; Fib_collagen_C.	RA	"Covalent structure of collagen: amino acid sequence of cyanogen bromide peptides from the amino-terminal segment of type III collagen of human liver.";
DR	InterPro; IPR001007; WFEC.	RT	Nucleic Acids Res. 17:6742-6742(1989).
DR	InterPro; IPR0191; Collagen; 13.	RN	[3]
DR	Probdom; PD002078; Fib_collagen_C; 1.	RP	SEQUENCE OF 168-398.
DR	SMART; SM00214; WFEC; 1.	RX	Medline=77134724; Pubmed=557335;
DR	PROSITE; PS01208; WFEC; 1.	RA	Seyer J.M., Kang A.H.;
KW	Extracellular matrix; Connective tissue; Repeat; Hydroxylation;	RA	"Covalent structure of collagen: amino acid sequence of the entire human alpha 1 (III) collagen.";
KW	Glycoprotein; Collagen; Signal.	RT	Biochemistry 16:1158-1164(1977).
FT	SIGNAL 1 23 POTENTIAL.	RN	[4]
PROPER	24 144 AMINO-TERMINAL PROPEPTIDE (BY SIMILARITY).	RP	REVISIONS.
FT	CHAIN 145 1003 COLLAGEN ALPHA 1(III) CHAIN.	RN	Seyer J.M.;
FT	PROPEP 1004 1262 CARBOXYL-TERMINAL PROTEIPE (BY SIMILARITY).	RP	Submitted (DEC-1977) to the PIR data bank.
FT	DOMAIN 29 88 VWF.	RN	[5]
FT	DOMAIN 145 164 NONHELICAL REGION (N-TERMINAL) (BY SIMILARITY).	RP	SEQUENCE OF 399-727.
FT	DOMAIN 165 994 TRIPLE-HELICAL REGION (BY SIMILARITY).	RX	Medline=79000343; Pubmed=687591;
FT	DOMAIN 995 1003 NONHELICAL REGION (C-TERMINAL) (BY SIMILARITY).	RA	Seyer J.M., Kang A.H.;
FT	NON_CONS 886 887 INTERCHAIN (BY SIMILARITY).	RT	"Covalent structure of collagen: amino acid sequence of five consecutive C-terminal peptides from type III collagen of human liver.";
FT	NON_CONS 922 923 INTERCHAIN (BY SIMILARITY).	RT	Biochemistry 17:3404-3411(1978).
FT	DISULFD 994 994 INTERCHAIN (BY SIMILARITY).	RN	[6]
FT	DISULFD 995 995 HYDROXYLATION (BY SIMILARITY).	RP	SEQUENCE OF 728-964.
FT	MOD_RES 262 262 HYDROXYLATION (BY SIMILARITY).	RX	Medline=80198282; Pubmed=6246925;
FT	MOD_RES 283 283 HYDROXYLATION (BY SIMILARITY).	RA	Seyer J.M., Mainardi C., Kang A.H.;
FT	MOD_RES 859 859 HYDROXYLATION (BY SIMILARITY).	RT	"Covalent structure of collagen: amino acid sequence of alpha 1 (III)-CB5 from type III collagen of human liver.";
FT	CARBONYD 1163 1163 N-LINKED (GLCNAC. . .) (POTENTIAL).	RT	Biochemistry 19:1583-1589(1980).
FT	CONFLICT 96 96 E->K (IN REF. 2).	RN	[7]
FT	CONFLICT 1132 1132 F->S (IN REF. 3).	RP	SEQUENCE OF 950-1466 FROM N.A.
SQ	SEQUENCE 1262 AA; 121249 MW; 96ABE7B2E9DEB43D CRC64;	RX	Medline=88169827; Pubmed=3357782;
Query Match Best Local Similarity 65.2%; Score 60; DB 1; Length 1262; Matches 10; conservative 3; Mismatches 1; Indels 0; Gaps 0;	RA	Mankoo B.S., Daigleish R.;	
QY 1 GVGDKGNPGWPCA 14	RT	"Human pro alpha 1(III) collagen: cDNA sequence for the 3' end.";	
DB 857 GVGERGSPGGCA 870	RT	Nucleic Acids Res. 16:2337-2337(1988).	
RESULT 18 CA13_HUMAN	RT	REVISION TO 1184.	
CA13_HUMAN	RT	Medline=8908346; Pubmed=3211760;	
ID CA13_HUMAN	RT	Molneux K., Daigleish R.;	
AC P02461; Q15112; DT 21-JUL-1986 (Rel. 01, Created) 01-JAN-1990 (Rel. 13, Last sequence update)	RL	"Human type III collagen 'variant' is a cDNA cloning artefact.";	
AC P02461; Q15112; DT 21-JUL-1986 (Rel. 01, Created) 01-JAN-1990 (Rel. 13, Last sequence update)	RN	Nucleic Acids Res. 16:11833-11833(1988).	
AC P02461; Q15112; DT 21-JUL-1986 (Rel. 01, Created) 01-JAN-1990 (Rel. 13, Last sequence update)	RP	"Molecular cloning and carboxyl-propeptide analysis of human type III procollagen.";	
AC P02461; Q15112; DT 21-JUL-1986 (Rel. 01, Created) 01-JAN-1990 (Rel. 13, Last sequence update)	RL	Nucleic Acids Res. 12:9383-9394(1984).	
AC P02461; Q15112; DT 21-JUL-1986 (Rel. 01, Created) 01-JAN-1990 (Rel. 13, Last sequence update)	RN	[10]	
AC P02461; Q15112; DT 21-JUL-1986 (Rel. 01, Created) 01-JAN-1990 (Rel. 13, Last sequence update)	RP	SEQUENCE OF 965-1200.	
AC P02461; Q15112; DT 21-JUL-1986 (Rel. 01, Created) 01-JAN-1990 (Rel. 13, Last sequence update)	RX	Medline=81208139; Pubmed=7016180;	

- RA Seyer J.M., Kang A.H.;  
 RT "Covalent structure of collagen: amino acid sequence of alpha  
 RT I (III)-CB9 from type III collagen of human liver.";  
 RL Biochemistry 20:2621-2627(1981).  
 RN [11]
- RN SEQUENCE OF 1176-1466 FROM N.A.  
 RX MEDLINE=85157600; PubMed=2579949;  
 RA Chu M.-L., Weil D., de Wet W.J., Bernard M.P., Sippola M., Ramirez F.;  
 RT "Isolation of cDNA and genomic clones encoding human pro-alpha<sub>1</sub><sup>1</sup> (III) collagen. Partial characterization of the 3' end region of the gene.";  
 RT J. Biol. Chem. 260:4357-4363(1985).  
 RL [12]
- RN SEQUENCE OF 1161-1200 FROM N.A.  
 RP MEDLINE=8618704; PubMed=3754462;  
 RA Miskulin M., Daigleish R., Kluwe-Beckerman B., Rennard S.I.,  
 RA Toloshev P., Brantly M., Crystal R.G.;  
 RT "Human type III collagen gene expression is coordinately modulated with the type I collagen genes during fibroblast growth.";  
 RL Biochemistry 25:1408-1413(1986).  
 RN [13]
- RC SEQUENCE OF 1-170 FROM N.A.  
 RX TISSUE=Placenta;  
 MEDLINE=88303360; PubMed=3405773;  
 RA Toman D., Ricca G., de Crombrugge B.;  
 RT "Nucleotide sequence of a cDNA coding for the amino-terminal region of human prepro alpha 1(III) collagen.";  
 RL Nucleic acids Res. 16:7201-7201(1988).  
 RN [14]
- RN SEQUENCE OF 1-176 FROM N.A.  
 RX MEDLINE=9378752; PubMed=2777083;  
 RA Benson-Chanda V., Su M.W., Weil D., Chu M.-L., Ramirez F.;  
 RT "Cloning and analysis of the 5' portion of the human type-III procollagen gene (COL3A1).";  
 RL Gene 78:255-265(1999).  
 RN [15]
- RP REVIEW ON VARIANTs.  
 RX MEDLINE=9725559; PubMed=9101290;  
 RA Kuivaniemi H., Tromp G., Prockop D.J.;  
 RT "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-associated collagen (type X), and network-forming collagen (type X) cause a spectrum of diseases of bone, cartilage, and blood vessels.";  
 RL Hum. Mutat. 9:300-315(1997).  
 RN [16]
- RP VARIANT AORTIC ANEURYSM ARG-303, AND VARIANT THR-668.  
 RX MEDLINE=9329388; PubMed=8514866;  
 RA Tromp G., Wu Y., Prockop D.J., Madhatheri S.L., Kleinert C., Abbott W.M.,  
 RA Earley J.J., Zhuang J., Noergaard O., Darling R.C., Abbott W.M.,  
 RA Cole C.W., Jaakkola P., Ryynenen M., Pearce W.H., Yao J.S.T.,  
 RA Majamaa K., Smullen S.N., Galatica Z., Ferrell R.E., Jimenez S.A.,  
 RA Jackson C.E., Michaels V.V., Kaye M., Kuivaniemi H.;  
 RT "Sequencing of cDNA from 50 unrelated patients reveals that mutations in the triple-helical domain of type III procollagen are an infrequent cause of aortic aneurysms.";  
 RL J. Clin. Invest. 91:2539-2545(1993).  
 RN [17]
- RP VARIANT THR-698.  
 RX MEDLINE=91045136; PubMed=2235526;  
 RA Zafarullah K., Klempert C., Tromp G., Kuivaniemi H., Kontusaari S.,  
 RA Wu Y., Ganguly A., Prockop D.J.;  
 RT "G to A polymorphism in exon 31 of the COL3A1 gene.";  
 RL Nucleic Acids Res. 18:6180-6180(1990).  
 RN [18]
- RP VARIANT AORTIC ANEURYSM ARG-786.  
 RX MEDLINE=91056145; PubMed=2241125;  
 RA Kontusaari S., Tromp G., Kuivaniemi H., Romanic A.M., Prockop D.J.;  
 RT "A mutation in the gene for type III procollagen (COL3A1) in a family with aortic aneurysms.";  
 RL J. Clin. Invest. 86:1465-1473(1990).  
 RN [19]
- RP VARIANT EDS-IV ARG-828.  
 RX MEDLINE=94016385; PubMed=8411057;  
 Richards A.J., Narcisi P., Lloyd J.C., Ferguson C., Pope F.M.;
- RT "The substitution of glycine 661 by arginine in type III collagen produces mutant molecules with different thermal stabilities and causes Ehlers-Danlos syndrome type IV.";  
 RT J. Med. Genet. 30:690-693(1993).  
 RN [20]
- RP VARIANT EDS-IV SER-957.  
 RX MEDLINE=89109135; PubMed=2492273;  
 RA Tromp G., Kuivaniemi H., Shikata H., Prockop D.J.;  
 RT "A single base mutation that substitutes serine for glycine 790 of the alpha 1 (III) chain of type III procollagen exposes an arginine and causes Ehlers-Danlos syndrome type IV.";  
 RT J. Biol. Chem. 264:1349-1352(1989).  
 RL [21]
- RN RP VARIANT EDS-IV VAL-960.  
 RX MEDLINE=952684229; PubMed=779417;  
 RA Tromp G., de Paep A., Nyutinck L., Madhatheri S.L., Kuivaniemi H.;  
 RT "Substitution of valine for glycine 793 in type III procollagen in Ehlers-Danlos syndrome type IV.";  
 RL Hum. Mutat. 5:179-181(1995).  
 RN [22]
- RP VARIANT EDS-IV GLU-1014.  
 RX MEDLINE=92316511; PubMed=1352273;  
 RA Richards A.J., Ward P.N., Narcisi P., Nicholls A.C., Lloyd J.C.,  
 RA Pope F.M.;  
 RT "A single base mutation in the gene for type III collagen (COL3A1) converts glycine 847 to glutamic acid in a family with Ehlers-Danlos syndrome type IV. An unaffected family member is mosaic for the mutation.";  
 RT RL Hum. Genet. 89:414-418(1992).  
 RL [23]
- RP VARIANT EDS-IV ASP-1050.  
 RX MEDLINE=90037070; PubMed=2808425;  
 RA Tromp G., Kuivaniemi H., Stolle C.A., Pope F.M., Prockop D.J.;  
 RT "Single base mutation in the type III procollagen gene that converts the codon for glycine 883 to aspartate in a mild variant of Ehlers-Danlos syndrome type IV.";  
 RT RL J. Biol. Chem. 264:19313-19317(1989).  
 RL [24]
- RP VARIANT EDS-IV VAL-1077.  
 RX MEDLINE=91374480; PubMed=1895316;  
 RA Richards A.J., Lloyd J.C., Ward P.N., de Paep A., Narcisi P.,  
 RA Pope F.M.;
- RT "Characterisation of a glycine to valine substitution at amino acid position 910 of the triple helical region of type III collagen in a patient with Ehlers-Danlos syndrome type IV.";  
 RT RL J. Med. Genet. 28:458-463(1991).  
 RN [25]
- RP VARIANT EDS-IV GLU-1173.  
 RX MEDLINE=93022543; PubMed=1357232;  
 RA Johnson P.H., Richards A.J., Pope F.M., Hopkinson D.A.;  
 RA Johnson P.H., Richards A.J., Pope F.M.,  
 RESULT 19  
 CA34\_BOVIN  
 ID CA34\_BOVIN  
 AC STANDARD  
 AC 028084;  
 DT 01-Nov-1997 ( Rel. 35, created)  
 DT 01-Nov-1997 ( Rel. 35, last sequence update)  
 DT 01-Nov-1997 ( Rel. 35, last annotation update)  
 DE Collagen alpha 3 (IV) chain (Fragment).  
 DE COL4A3.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Butheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;  
 OC Bovidae; Bovinae; Bos.

		NCBI_TaxID=9913;
RN	[1]	SEQUENCE FROM N.A.
RP		
RT		TISSUE-Lens;
RX		MEDLINE=91093146; PubMed=1985905;
RA		Morrison K.E., Germine G.G., Reeder S.T.,
RA		'use of the polymerase chain reaction to clone and sequence a cDNA
RT		encoding the bovine alpha 3 chain of type IV collagen.';
RL		J. Biol. Chem. 266:34-39(1991).
RN	[2]	SEQUENCE OF 227-258.
RP		TISSUE-Kidney;
RX		MEDLINE=90202779; PubMed=2318822;
RA		Gunawardene S., Saus J., Noelken M.E., Hudson B.G.;
RT		'Glomerular basement membrane. Identification of a fourth chain,
RT		alpha 4, of type IV collagen.';
RL		J. Biol. Chem. 265:5466-5469(1990).
RN	[3]	SEQUENCE OF 227-254.
RP		SEQUENCE OF 227-244.
RX		MEDLINE=88330844; PubMed=417661;
RA		Saus J., Wieslander J., Langeveld J.P.M., Quiñones S., Hudson B.G.;
RT		'Identification of the Goodpasture antigen as the alpha 3(IV) chain
RT		of collagen IV.';
RL		J. Biol. Chem. 263:13374-13380(1988).
RN	[4]	SEQUENCE OF 227-244.
RP		MEDLINE=8722419; PubMed=2438283;
RX		Butkowska R.J., Langeveld J.P.M., Wieslander J., Hamilton J.,
RA		Hudson B.G.;
RT		'Localization of the goodpasture epitope to a novel chain of basement
RT		membrane collagen.';
RL		J. Biol. Chem. 262:7874-7877(1987).
CC		-I- FUNCTION: TYPE IV COLLAGEN IS THE MAJOR STRUCTURAL COMPONENT OF
CC		GLOMERULAR BASEMENT MEMBRANES (GEM), FORMING A 'CHICKEN-WIRE'
CC		NITROGEN NETWORK TOGETHER WITH LAMININS, PROTOGLYCAN AND ENTACTIN/
CC		-I- SUBUNIT: THERE ARE SIX TYPE IV COLLAGEN ISOFORMS, ALPHA 1(IV) -
CC		ALPHA 6(IV), EACH OF WHICH CAN FORM A TRIPLE HELIX STRUCTURE
CC		WITH 2 OTHER CHAINS IN A TRIPLE-Helix, AND A SHORT N-TERMINAL
CC		-I- SURCELLULAR LOCATION: CELL SURFACE (POTENTIAL).
CC		-I- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS
CC		G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY
CC		CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL
CC		TRIPLE-HELICAL 7S DOMAIN.
CC		-I- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPETIDE REPEATING
CC		UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC		-I- DOMAIN: TYPE IV COLLAGENS CONTAIN NUMEROUS CYSTEINE RESIDUES WHICH
CC		ARE INVOLVED IN INTER- AND INTRAMOLECULAR DISULFIDE BONDING. 12 OF
CC		THESE, LOCATED IN THE NCI DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE
CC		IV COLLAGENS.
CC		-I- SIMILARITY: TO OTHER TYPE IV COLLAGENS.
CC		This SWISS-PROT entry is copyright. It is produced through a collaboration
CC		between the European Bioinformatics Institute and the EMBL outstation -
CC		the European Bioinformatics Institute. There are no restrictions on its
CC		use by non-profit institutions as long as its content is in no way
CC		modified and this statement is not removed. Usage by and for commercial
CC		entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to license@isb-sib.ch).
CC		--
DR		EMBL; M63139; AAH62708.1; -.
DR		InterPro; IPR01442; C4.
DR		InterPro; IPR000887; Collagen.
DR		Pfam; PF01413; C4; 2.
DR		Pfam; PF0191; Collagen; 4.
DR		ProDom; PD003923; C4; 2.
DR		SMART; SM00111; C4; 2.
DR		Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW		Glycoprotein; Basement membrane; Collagen; Cell adhesion.
FT		NON_TER 1
FT		DOMAIN 1 238
FT		DOMAIN 239 471
FT		NONHELCICAL REGION (NCI).
FT		
FT		SITE 106 108
FT		MOD_RES 232 232
FT		MOD_RES 238 238
FT		DISULFID 261 352
FT		DISULFID 294 349
FT		DISULFID 306 312
FT		DISULFID 371 466
FT		DISULFID 405 463
FT		DISULFID 417 423
FT		DISULFID 253 S -> Y (IN REF. 3).
FT		SEQUENCE 471 AA: 47585 MW: C03B66F1487008DE CRC64;
FT		
FT		Query Match 64.1%; Score 59; DB 1; Length 471;
FT		Best Local Similarity 66.7%; Pred. No. 0; 13; Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
FT		Db 116 GRKGKGKGSFGFP 130
FT		
FT		RESULT 20
FT		CAL4_CAEEL STANDARD; PRT; 1758 AA.
FT		ID CA14_CAEEL
FT		AC P17139;
FT		DT 01-AUG-1990 (Rel. 15. Created)
FT		DT 01-MAR-1992 (Rel. 21. Last sequence update)
FT		DT 01-MAR-2002 (Rel. 41. Last annotation update)
FT		DE Collagen alpha 1(IV) chain precursor.
FT		GN EXP-9 OR CLB-2 OR KO444.1.
FT		OS Caenorhabditis elegans.
FT		OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
FT		OC Nematidae; Plecterinae; Caenorhabditis;
FT		OX NCBI_TaxID=6239;
FT		
FT		RP SEQUENCE FROM N.A.
FT		RX MEDLINE=9111582; PubMed=1996137;
FT		RA Guo X., Johnson J., Kramer J.M.;
FT		'Embryonic lethality caused by mutations in basement membrane
FT		collagen of C. elegans.';
FT		RL Nature 349:707-709(1991).
FT		
FT		RP SEQUENCE FROM N.A.
FT		RX MEDLINE=9415018; PubMed=7906398;
FT		RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
FT		RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
FT		RA Craxton M., Dear S., Du Z., Durbin R., Favell A., Fraser A.,
FT		RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
FT		RA Johnston L., Jones M., Kershaw J., Kirsten J., Lalisse N.,
FT		RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
FT		RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkeen R.,
FT		RA Sims M., Smaldon N., Smith A., Smith M., Sonnhammer E., Staden R.,
FT		RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
FT		RA Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,
FT		RA Wohldman P.;
FT		RT 2.2 Mo of contiguous nucleotide sequence from chromosome III of C. elegans. ";
FT		RL Nature 368:32-38(1994).
FT		
FT		RP SEQUENCE OF 1446-1758 FROM N.A.
FT		RX MEDLINE=9008929; PubMed=2793871;
FT		RA Guo X., Kramer J.M.;
FT		'The two Caenorhabditis elegans basement membrane (type IV) collagen genes are located on separate chromosomes.';
FT		J. Biol. Chem. 264:17574-17582(1989).
FT		CC -I- FUNCTION: COLLAGEN TYPE IV IS SPECIFIC FOR BASEMENT MEMBRANES.
FT		-I- SUBUNIT: TRIMERS OF TWO ALPHA 1(IV) AND ONE ALPHA 2(IV) CHAIN.
FT		TYPE IV COLLAGEN FORMS A MESH-LIKE NETWORK LINKED THROUGH
FT		INTERMOLECULAR INTERACTIONS BETWEEN 7S DOMAINS AND BETWEEN NCI1



RT	collagen and tissue-specific differences in the expression of the corresponding transcripts.";	CC	detachment, macular abnormalities and occipital encephalocele.
RN	MATRIX Biol. 16:319-328(1998).	CC	
[2]	SEQUENCE FROM N.A.	CC	
RX	MEDLINE-20289799; PubMed=10830953;	CC	
RA	Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T., Soeda E., Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Horisieck G., Hornischer K., Bleschmidt K., Polley A., Oki M., Takegi T., Saito K., Taudien S., Blechschmidt K., Dehm M., Menzel U., Delabar J., Kumpf K., Dehm M., Patterson D., Reichwald K., Rump J., Schillhabel M., Schudy A., Zimmermann W., Roseman A., Kudo J., Shibusawa K., Kawasaki K., Asakawa S., Shintani A., Sasaki T., Nagamine K., Mitsuhashi S., Antonarakis S.E., Minoshima S., Shimizu N., Nordsiek G., Reichelt J., Kauer G., Bloechert H., Scherzer M., Schoen O., Desario A., Dehm M., Lammel L., Dagand E., Reimann J., Beck A., Klages S., Heinig S., Riesseleman L., Wehrmeyer S., Borzym K., Gardner K., Nizetic D., Francis F., Lehrach H., Reinhardt R., Yapo M.-L., "The DNA sequence of human chromosome 21.", Nature 405:311-319(2000).	CC	
RN	NATURE 405:311-319(2000).	CC	
RX	SEQUENCE OF 834-1516 FROM N.A.	CC	
RA	MEDLINE-94245237; PubMed=8188291; Oh S.P., Warmann M.L., Seldin M.F., Cheng S., Knoll J.H., Timmons S., Olsen B.R.; "Cloning of cDNA and genomic DNA encoding human type XVIII collagen and localization of the alpha 1(XVIII) collagen gene to mouse chromosome 10 and human chromosome 21.", Genomics 19:494-499(1994).	CC	
RN	[4] SEQUENCE OF 1334-1516 FROM N.A.	CC	
RC	"Tissue=placenta;"	CC	
RA	Zhi-Tong H., Biao L., Wei-Jie Z., Xiang-Fu W.; "Cloning and expression of human endostatin gene in Escherichia coli.", Submitted (SEPP-1999) to the EMBL/GenBank/DBJ databases.	CC	
RT	[5] INVOLVEMENT IN KNOBLICH SYNDROME.	CC	
RX	MEDLINE-20404145; PubMed=10942431; Seftie A.L., Sossi V., Camargo A., Zatz M., Brahe C., Passos-Bueno M.R.; "Collagen XVII, containing an endogenous inhibitor of angiogenesis and tumor growth, plays a critical role in the maintenance of retinal structure and in neural tube closure.", Hum. Mol. Genet. 9:2051-2058(2000).	CC	
RN	[6] VARIANT ASN-1437.	CC	
RX	PUBMED=11606364; Iughetti P., Suzuki O., Godoi P.H., Alves V.A., Seftie A.L., Zorick T., Soares F., Camargo A., Moreira E.S., di Loreto C., Moreira-Filho C.A., Simpson A., Oliva G., Passos-Bueno M.R.; "A polymorphism in endostatin, an angiogenesis inhibitor, predisposes for the development of prostatic adenocarcinoma.", Cancer Res. 61:7315-7318(2001).	CC	
RL	-1- FUNCTION: COL18A1 PROBABLY PLAYS A MAJOR ROLE IN DETERMINING THE RETINAL STRUCTURE AS WELL AS IN THE CLOSURE OF THE NEURAL TUBE.	CC	
RX	PUBMED=11606364; Iughetti P., Suzuki O., Godoi P.H., Alves V.A., Seftie A.L., Zorick T., Soares F., Camargo A., Moreira E.S., di Loreto C., Moreira-Filho C.A., Simpson A., Oliva G., Passos-Bueno M.R.; "A polymorphism in endostatin, an angiogenesis inhibitor, predisposes for the development of prostatic adenocarcinoma.", Cancer Res. 61:7315-7318(2001).	CC	
RT	-1- FUNCTION: ENDOSTATIN POTENTLY INHIBITS ENDOTHELIAL CELL PROLIFERATION AND ANGIOGENESIS. MAY INHIBIT ANGIOGENESIS BY BINDING TO THE HEPARAN SULPHATE PROTEOGLYCANS INVOLVED IN GROWTH FACTOR SIGNALLING.	CC	
CC	-1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A SHORT FORM/NCI-303 AND A LONG FORM/NCI-493 (SHOWN HERE); ARE PRODUCED BY ALTERNATIVE SPlicing.	CC	
CC	-1- TISSUE SPECIFICITY: PRESENT IN MULTIPLE ORGANS WITH HIGHEST LEVELS IN LIVER, LUNG AND KIDNEY.	CC	
CC	-1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPETIDE REPEATING UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.	CC	
CC	-1- POLYMORPHISM: There is an association between a polymorphism in position 1437 and prostate cancer. Heterozygous Asn-1437 individuals have a 2.5 times increased chance of developing prostate cancer as compared with homozygous Asp-1437 individuals.	CC	
CC	-1- DISEASE: Defects in COL18A1 are a cause of Knobloch syndrome (KS or KNO); an autosomal recessive disorder defined by the occurrence of high myopia, vitreoretinal degeneration with retinal	CC	
CC	[1] -1- SIMILARITY: BELONGS TO THE FIBRIL-ASSOCIATED COLLAGENS WITH INTERRUPTED HELICES (FACIT) FAMILY.	CC	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation of the European Bioinformatics Institute. There are no restrictions on use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage for commercial entities requires a license agreement (see <a href="http://www.isb-sib.ch/announcements/">http://www.isb-sib.ch/announcements/</a> or send an email to license@isb-sib.ch).	CC	
CC	use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage for commercial entities requires a license agreement (see <a href="http://www.isb-sib.ch/announcements/">http://www.isb-sib.ch/announcements/</a> or send an email to license@isb-sib.ch).	CC	
CC	EMBL; AF018081; AAC39658.1; -;	CC	
DR	EMBL; AF018082; AAC39659.1; -;	CC	
DR	EMBL; AL163302; CAB00482.1; -;	CC	
DR	EMBL; L22548; AA51864.1; -;	CC	
DR	EMBL; AF180460; AAF01310.1; ALT_INIT;	CC	
DR	HSSP; P3961; IKG;	CC	
DR	GlycositDB; P30060; -;	CC	
DR	MIM; 120328; -;	CC	
DR	MIM; 267750; -;	CC	
DR	InterPro; IPR000887; Collagen;	CC	
DR	InterPro; IPR001791; Laminin_G;	CC	
DR	InterPro; IPR00129; TSPN;	CC	
DR	PFAM; PF02210; TSPN; 1;	CC	
DR	SMART; SM00282; LamG; 1;	CC	
DR	SMART; SM00210; TSPN; 1;	CC	
KW	Extracellular matrix; Connective tissue; Repeat; Hydroxylation; Cell adhesion; Collagen; Glycoprotein; Signal; Alternative splicing; Polymorphism.	CC	
KW	SIGNAL 1 23 POTENTIAL.	CC	
FT	CHAIN 24 1516 COLLAGEN ALPHA 1(XVIII) CHAIN.	CC	
FT	CHAIN 1334 1516 ENDOSTATIN.	CC	
FT	DOMAIN 24 516 NONHELICAL REGION 1 (NC1).	CC	
FT	DOMAIN 517 550 TRIPLE-HELICAL REGION 1 (COL1).	CC	
FT	DOMAIN 551 560 NONHELICAL REGION 2 (NC2).	CC	
FT	DOMAIN 561 640 TRIPLE-HELICAL REGION 2 (COL2).	CC	
FT	DOMAIN 641 664 NONHELICAL REGION 3 (NC3).	CC	
FT	DOMAIN 665 786 TRIPLE-HELICAL REGION 3 (COL3).	CC	
FT	DOMAIN 787 809 NONHELICAL REGION 4 (NC4).	CC	
FT	DOMAIN 810 892 TRIPLE-HELICAL REGION 4 (COL4).	CC	
FT	DOMAIN 893 906 NONHELICAL REGION 5 (NC5).	CC	
FT	DOMAIN 907 948 TRIPLE-HELICAL REGION 5 (COL5).	CC	
FT	DOMAIN 949 961 NONHELICAL REGION 6 (NC6).	CC	
FT	DOMAIN 962 1034 TRIPLE-HELICAL REGION 6 (COL6).	CC	
FT	DOMAIN 1035 1044 NONHELICAL REGION 7 (NC7).	CC	
FT	DOMAIN 1045 1077 TRIPLE-HELICAL REGION 7 (COL7).	CC	
FT	DOMAIN 1078 1089 NONHELICAL REGION 8 (NC8).	CC	
FT	DOMAIN 1090 1111 TRIPLE-HELICAL REGION 8 (COL8).	CC	
FT	DOMAIN 1112 1118 NONHELICAL REGION 9 (NC9).	CC	
FT	DOMAIN 1119 1173 TRIPLE-HELICAL REGION 9 (COL9).	CC	
FT	DOMAIN 1174 1186 NONHELICAL REGION 10 (NC10).	CC	
FT	DOMAIN 1187 1204 TRIPLE-HELICAL REGION 10 (COL10).	CC	
FT	DOMAIN 1205 1516 NONHELICAL REGION 11 (NC11).	CC	
FT	CARBOHYD 68 68 N-LINKED (GLCNAC . . . ) (POTENTIAL).	CC	
FT	CARBOHYD 129 129 N-LINKED (GLCNAC . . . ) (POTENTIAL).	CC	
FT	CARBOHYD 164 164 N-LINKED (GLCNAC . . . ) (POTENTIAL).	CC	
FT	CARBOHYD 691 691 N-LINKED (GLCNAC . . . ) (POTENTIAL).	CC	
FT	CARBOHYD 1329 1329 O-LINKED (GALNAc . . . ).	CC	
FT	/FTG-CARL_000150. SHORT ISOFORM.	CC	
FT	D > N.	CC	
FT	VARIANT 1437 1437 /FTGd-VAR_012709.	CC	
FT	F -> S (IN REF. 2).	CC	
FT	I -> V (IN REF. 2).	CC	
FT	V -> L (IN REF. 3).	CC	

CC	G-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL TRIPLE-HELICAL TS DOMAIN.
CC	-!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPPELEXTENDING UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS. THESE, LOCATED IN THE NCI DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE IV COLLAGENS.
CC	-!- DISEASE: MUTATIONS IN LEP-2 ARE GENERALLY EMBRYONIC LETHAL.
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> ) or send an email to license@isb-sib.ch).
CC	-----
QY	1 GVKGDKGNFGWPGAP 15
Db	517 GLKGOKGEPSVPGPP 531
RESULT 23	
CA24_CAEEL	
ID	
AC	P17140; P17140;
DT	01-AUG-1994 (Rel. 15, Created)
DT	01-OCT-1994 (Rel. 30, Last sequence update)
DT	30-MAY-2000 (Rel. 39, Last annotation update)
DE	Collagen alpha 2(IV) chain precursor.
GN	LET-2 OR CIB-1.
OS	Caenorhabditis elegans.
OC	Bukay-Yoav; Metzoco; Nematoda; Chromadorea; Rhabditoidea; Rhabditida; Metazoa; Nemata; Chromadorea; Rhabditoidea; Rhabditidae; Peltorhinidae; Caenorhabditis.
OX	NCBI_TaxID=6239;
[1]	SEQUENCE FROM N.A.
RP	STRAIN=BRISTOL_N2;
RX	MEDLINE=94012964; PubMed=7691828;
RA	Sibley M.H., Johnson J.J., Mello C.C., Kramer J.M.;
RT	*Genetic identification, sequence, and alternative splicing of the Caenorhabditis elegans alpha 2(IV) collagen gene.;
RL	J. Cell Biol. 123:255-264(1993).
RN	[2]
RP	PRELIMINARY SEQUENCE OF 1495-1758 FROM N.A.
RC	STRAIN=BRISTOL_N2;
RX	MEDLINE=9008929; PubMed=2793871;
RA	Guo X., Kramer J.M.;
RT	"The two Caenorhabditis elegans basement membrane (type IV) collagen genes are located on separate chromosomes.;"
RL	J. Biol. Chem. 264:17574-17582(1989).
RN	[3]
RP	VARIANTS.
RX	MEDLINE=94320591; PubMed=8045258;
RA	Sibley M.H., Graham P.L., von Mende N., Kramer J.M.;
RT	*Mutations in the alpha 2(IV) basement membrane collagen gene of Caenorhabditis elegans produce phenotypes of differing severities.;"
RT	EMBO J. 13:3278-3285(1994).
RL	"FUNCTION: COLLAGEN TYPE IV IS SPECIFIC FOR BASEMENT MEMBRANES. SUBUNIT: TRIMERS OF TWO ALPHA 1(IV) AND ONE ALPHA 2(IV) CHAIN. TYPE IV COLLAGEN FORMS A MESH-LIKE NETWORK LINKED THROUGH INTERMOLECULAR INTERACTIONS BETWEEN 7S DOMAINS AND BETWEEN NCI DOMAINS."
CC	-!- ALTERNATIVE PRODUCTS: 2 ISOFORMS: I (SHOWN HERE) AND II; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC	-!- TISSUE SPECIFICITY: FORM I IS PREDOMINANT IN EMBRYOS AND FORM II IS PREDOMINANT IN THE LARVAE AND ADULTS.
CC	-!- DOMAIN: ALPHA CHAINS OF TYPE IV COLLAGEN HAVE A NONCOLLAGENOUS DOMAIN (NCL) AT THEIR C-TERMINUS, FREQUENT INTERRUPTIONS OF THE
CC	6-X-Y REPEATS IN THE LONG CENTRAL TRIPLE-HELICAL DOMAIN (WHICH MAY CAUSE FLEXIBILITY IN THE TRIPLE HELIX), AND A SHORT N-TERMINAL TRIPLE-HELICAL TS DOMAIN.
CC	-!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPPELEXTENDING UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS. THESE, LOCATED IN THE NCI DOMAIN, ARE CONSERVED IN ALL KNOWN TYPE IV COLLAGENS.
CC	-!- DISEASE: MUTATIONS IN LEP-2 ARE GENERALLY EMBRYONIC LETHAL.
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> ) or send an email to license@isb-sib.ch).
CC	-----
QY	1 GVKGDKGNFGWPGAP 15
Db	416 GPKGDGGDGPGAP 430
Query Match	63.0%; Score 58; DB 1; Length 1516; Best Local Similarity 66.7%; Pred. No. 0.57; 1; Mismatches 4; Indels 0; Gaps 0;
Match	11; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Query Match	63.0%; Score 58; DB 1; Length 1758; Best Local Similarity 73.3%; Pred. No. 0.66; 1; Mismatches 3; Indels 0; Gaps 0;
Match	11; Conservative 1; Mismatches 1; Indels 0; Gaps 0;



RT "Complete primary structure of rainbow trout type I collagen  
RT consisting of alpha<sub>1</sub>(I)alpha<sub>2</sub>(I)alpha<sub>3</sub>(I) heterotrimers.";  
RL Eur. J. Biochem. 268:2817-2827(2001).  
RN [2]  
SEQUENCE OF 417-1356 FROM N.A.  
RC TISSUE=FBroblast;  
RA Saito M., Kunisaki N., Hiroto I., Aoki T., Ishida M., Urano N.,  
Kimura S.;  
RT "Partial characterization of cDNA clones encoding the three distinct  
pro alpha chains of type I collagen from rainbow trout.";  
RL Fisheries Sci. 64:785-798.  
CC -I- FUNCTION: TYPE I COLLAGEN IS A MEMBER OF GROUP I COLLAGEN  
(FIBRILLAR FORMING COLLAGEN).  
CC -I- SUBUNIT: TRIMERS OF ONE ALPHA 2(I) AND TWO ALPHA 1(I) CHAINS.  
CC -I- TISSUE SPECIFICITY: FORMS THE FIBRILLS OF TENDON, LIGAMENTS AND  
BONES. IN BONES THE FIBRILLS ARE MINERALIZED WITH CALCIUM  
CC -I- PROLINES AT THE THIRD POSITION OF THE TRipeptide REPEATING  
UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
or send an email to license@isb-sib.ch).  
CC  
CC EMBL; AB052837; BAB5663\_1;  
DR EMBL; AB008372; BA33379\_1;  
DR Interpro; IPR000887; Collagen.  
DR Pfam; PF01410; COFL; 1.  
DR Pfam; PF01391; Collagen; 17.  
DR PRODOM; PDD02078; Fib\_collagen\_C; 1.  
DR SMART; SM0038; COLFI; 1.  
DR KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
KW Glycoprotein; Collagen; Signal;  
FT SIGNAL 1 24 POTENTIAL.  
FT PROPEP 25 ? AMINO TERMINAL PROPEPTIDE (POTENTIAL).  
FT CHAIN ? 1096 COLLAGEN ALPHA 2(I) CHAIN.  
FT PROPER 1097 1356 CARBOXYL-TERMINAL PROPEPTIDE  
(BY SIMILARITY).  
FT CARBOHYD 1257 1257 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 1356 AA; 126985 MW; 7B2F1F80DB10C93 CRC64;

Query Match 60.9%; Score 57; DB 1; Length 1356;  
Best Local Similarity 71.4%; Pred. No. 0.71; Matches 10; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Oy 1 GVKGDKGNPGPGKA 14  
Db 937 GFKGDGRGPSPGA 950

---

RESULT 27

MTCO\_MOUSE STANDARD; PRT; 518 AA.

ID MTCO.MOUSE STANDARD; PRT; 518 AA.

AC Q67054; 067054; 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Macrophage receptor MARCO (Macrophage receptor with collagenous  
DE structure).  
GN MARCO.  
OS Mus musculus (mouse).  
OC Bokaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxId=10090;  
RN {1}  
RP SEQUENCE FROM N.A.  
RX MEDLINE=95171455; PubMed=787067;  
RA Blomaa O., Kangas M., Sahlberg C., Tuukkanen J., Sormunen R.,

RA Liakka A., Theisleff I., Kraal G., Tryggvason K.;  
RP Cloning of a novel bacteria-binding receptor structurally related to  
RT scavenger receptors and expressed in a subset of macrophages.";  
RL Cell 80:603-609(1995).  
RN [2]  
SEQUENCE FROM N.A.  
RP SBQUENCE FROM N.A.  
RX MEDLINE=99265975; PubMed=10331948;  
RA Kangas M., Brannstrom A., Elomaa O., Matsuda Y., Eddy R., Shows T.B.,  
RT Tryggvason K.;  
RL "Structure and chromosomal localization of the human and murine genes  
for the macrophage MARCO receptor.";  
RN Genomics 50:82-89(1999).  
CC -I- FUNCTION: Bind Gram-positive and Gram-negative bacteria.  
CC -I- SUBCELLULAR LOCATION: Type II membrane protein (Probable).  
CC -I- TISSUE SPECIFICITY: Expressed in subpopulations of macrophages in  
the spleen and the medullary cord of lymph nodes.  
CC -I- SIMILARITY: CONTAINS A COLLAGEN DOMAIN.  
CC -I- SIMILARITY: CONTAINS 1 SRCR DOMAIN.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
or send an email to license@isb-sib.ch).  
CC  
CC EMBL; U18424; AAA68638\_1;  
DR EMBL; AF128423; AAD51136\_1;  
DR EMBL; AF127927; AAD51136\_1; JOINED.  
DR EMBL; AF127928; AAD51136\_1; JOINED.  
DR EMBL; AF128169; AAD51136\_1; JOINED.  
DR EMBL; AF128170; AAD51136\_1; JOINED.  
DR EMBL; AF128171; AAD51136\_1; JOINED.  
DR EMBL; AF127601; AAD51136\_1; JOINED.  
DR EMBL; AF127602; AAD51136\_1; JOINED.  
DR EMBL; AF128419; AAD51136\_1; JOINED.  
DR EMBL; AF128420; AAD51136\_1; JOINED.  
DR EMBL; AF128421; AAD51136\_1; JOINED.  
DR EMBL; AF128422; AAD51136\_1; JOINED.  
DR MGD; MGI:1309998; Marco.  
DR Interpro; IPR000887; Collagen.  
DR Interpro; IPR001190; SRCR.  
DR Pfam; PF00530; SRCR; 1.  
DR PRINTS; PRO0258; SPERACTRCPTR.  
DR PROSITE; PS00420; SRCR\_1; FALSE\_NEG.  
DR PROSITE; PS50287; SRCR\_2; 1.  
FT DOMAIN 1 48 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 4 9 69 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)  
(POTENTIAL).  
FT DOMAIN 70 518 EXTRACELLULAR (POTENTIAL).  
FT DOMAIN 149 418 COLLAGEN-LIKE.  
FT DOMAIN 423 518 SRCR.  
FT DISULFID 446 507 BY SIMILARITY.  
FT DISULFID 459 517 BY SIMILARITY.  
FT DISULFID 487 497 BY SIMILARITY.  
FT CARBOHYD 87 87 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 138 138 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 518 AA; 52730 MW; B0957601ECA23637 CRC64;

Query Match 60.9%; Score 56; DB 1; Length 518;  
Best Local Similarity 66.7%; Pred. No. 0.39; Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Oy 1 GVKGDKGNPGPGKA 15  
Db 294 GVKGDGKGPVQGV 308

---

RESULT 28

CALI\_HUMAN  
 ID CALI\_HUMAN STANDARD; PRM; 1143 AA.  
 AC O14993; O13676; O12885; Q9H572; Q9NPZ2;  
 DT 01-MAR-2002 (Rel. 41; Created)  
 DT 01-MAR-2002 (Rel. 41; Last sequence update)  
 DT 01-MAR-2002 (Rel. 41; Last annotation update)  
 DE Collagen alpha 1(XIX) chain precursor (Collagen alpha 1(Y) chain).  
 GN COL19A1.  
 OC Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OC NCBI\_TaxID:9606;  
 RN [1] SEQUENCE FROM N.A.  
 RP MEDLINE:92523914; PubMed:7775380;  
 RA Inoguchi K.; Yoshioka H.; Khaleduzzaman M.; Ninomiya Y.;  
 RT "The mRNA for alpha 1(XIX) collagen chain, a new member of FACITs,  
 contains a long unusual 3' untranslated region and displays many  
 unique splicing variants.";  
 J. Biochem. 117:137-146(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE:98008918; PubMed:9344653;  
 RA Khaleduzzaman M.; Sumiyoshi H.; Ueki Y.; Inoguchi K.; Ninomiya Y.;  
 RA Yoshioka H.;  
 RT "Structure of the human type XIX collagen (COL19A1) gene, which  
 suggests it has arisen from an ancestor gene of the FACIT family.";  
 RL Genomics 45:304-312(1997).  
 RN [3]  
 RP SEQUENCE OF 132-952 FROM N.A.  
 RA Blakey S.; Parker A.; Skuce C.;  
 DR Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE OF 739-1143 FROM N.A.  
 RC TISSUE-SKIN;  
 RX MEDLINE:94308092; PubMed:8034603;  
 RA Myers J.C.; Yang H.; D'Ippolito J.A.; Presente A.; Miller M.K.;  
 DR Dion A.S.;  
 RT "The triple-helical region of human type XIX collagen consists of  
 multiple collagenous subdomains and exhibits limited sequence homology  
 to alpha 1(XVI).";  
 RT 269:18549-18557(1994).  
 CC -1- FUNCTION: May act as a cross-bridge between fibrils and other  
 extracellular matrix molecules.  
 CC -1- PTM: PROLINE AT THE THIRD POSITION OF THE TRIPETIDE REPEATING  
 UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS (BY  
 J. Biol. Chem. 269:18549-18557(1994)).  
 CC -1- SIMILARITY: BELONGS TO THE FIBRIL-ASSOCIATED COLLAGENS WITH  
 CC INTERRUPTED HELICES (FACIT) FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>  
 or send an email to license@isb-sib.ch).  
 CC -----  
 DR EMBL: D38163; BAA07368; 1; -  
 DR EMBL: AB004629; BAA23309; 1; -  
 DR EMBL: AB004634; BAA23309; 1; JOINED.  
 DR EMBL: AB004604; BAA23309; 1; JOINED.  
 DR EMBL: AB004615; BAA23309; 1; JOINED.  
 DR EMBL: AB004626; BAA23309; 1; JOINED.  
 DR EMBL: AB004630; BAA23309; 1; JOINED.  
 DR EMBL: AB004631; BAA23309; 1; JOINED.  
 DR EMBL: AB004632; BAA23309; 1; JOINED.  
 DR EMBL: AB004633; BAA23309; 1; JOINED.  
 DR EMBL: AB004634; BAA23309; 1; JOINED.  
 DR EMBL: AB004635; BAA23309; 1; JOINED.  
 DR EMBL: AB004636; BAA23309; 1; JOINED.  
 DR EMBL: AB004637; BAA23309; 1; JOINED.  
 DR EMBL: AB004638; BAA23309; 1; JOINED.  
 DR EMBL: AB004639; BAA23309; 1; JOINED.  
 DR EMBL: AB004640; BAA23309; 1; JOINED.  
 DR EMBL: AB004641; BAA23309; 1; JOINED.  
 DR EMBL: AB004642; BAA23309; 1; JOINED.  
 DR EMBL: AB004643; BAA23309; 1; JOINED.  
 DR EMBL: AB004644; BAA23309; 1; JOINED.  
 DR EMBL: AB004645; BAA23309; 1; JOINED.  
 DR EMBL: AB004646; BAA23309; 1; JOINED.  
 DR EMBL: AB004647; BAA23309; 1; JOINED.  
 DR EMBL: AB004648; BAA23309; 1; JOINED.  
 DR EMBL: AB004649; BAA23309; 1; JOINED.  
 DR EMBL: AB004650; BAA23309; 1; JOINED.  
 DR EMBL: AB004651; BAA23309; 1; JOINED.  
 DR EMBL: AB004652; BAA23309; 1; JOINED.  
 DR EMBL: AB004653; BAA23309; 1; JOINED.  
 DR EMBL: AB004654; BAA23309; 1; JOINED.  
 DR EMBL: AB004655; BAA23309; 1; JOINED.  
 DR EMBL: AB004656; BAA23309; 1; JOINED.  
 DR EMBL: AB004657; BAA23309; 1; JOINED.  
 DR EMBL: AB004658; BAA23309; 1; JOINED.  
 DR EMBL: AB004659; BAA23309; 1; JOINED.  
 DR EMBL: AB004660; BAA23309; 1; JOINED.  
 DR EMBL: AB004661; BAA23309; 1; JOINED.  
 DR EMBL: AB004662; BAA23309; 1; JOINED.  
 DR EMBL: AB004663; BAA23309; 1; JOINED.  
 DR EMBL: AB004664; BAA23309; 1; JOINED.  
 DR EMBL: AB004665; BAA23309; 1; JOINED.  
 DR EMBL: AB004666; BAA23309; 1; JOINED.  
 DR EMBL: AB004667; BAA23309; 1; JOINED.  
 DR EMBL: AB004668; BAA23309; 1; JOINED.  
 DR EMBL: AB004669; BAA23309; 1; JOINED.  
 DR EMBL: AB004670; BAA23309; 1; JOINED.  
 DR EMBL: AB004671; BAA23309; 1; JOINED.  
 DR EMBL: AB004672; BAA23309; 1; JOINED.  
 DR EMBL: AB004673; BAA23309; 1; JOINED.  
 DR EMBL: AB004674; BAA23309; 1; JOINED.  
 DR EMBL: AB004675; BAA23309; 1; JOINED.  
 DR EMBL: AB004676; BAA23309; 1; JOINED.  
 DR EMBL: AB004677; BAA23309; 1; JOINED.  
 DR EMBL: AB004678; BAA23309; 1; JOINED.  
 DR EMBL: AB004679; BAA23309; 1; JOINED.  
 DR EMBL: AB004680; BAA23309; 1; JOINED.  
 DR EMBL: AB004681; BAA23309; 1; JOINED.  
 DR EMBL: AB004682; BAA23309; 1; JOINED.  
 DR EMBL: AB004683; BAA23309; 1; JOINED.  
 DR EMBL: AB004684; BAA23309; 1; JOINED.  
 DR EMBL: AB004685; BAA23309; 1; JOINED.  
 DR EMBL: AB004686; BAA23309; 1; JOINED.  
 DR EMBL: AB004687; BAA23309; 1; JOINED.  
 DR EMBL: AB004688; BAA23309; 1; JOINED.  
 DR EMBL: AB004689; BAA23309; 1; JOINED.  
 DR EMBL: AB004690; BAA23309; 1; JOINED.  
 DR EMBL: AB004691; BAA23309; 1; JOINED.  
 DR EMBL: AB004692; BAA23309; 1; JOINED.  
 DR EMBL: AB004693; BAA23309; 1; JOINED.  
 DR EMBL: AB004694; BAA23309; 1; JOINED.  
 DR EMBL: AB004695; BAA23309; 1; JOINED.  
 DR EMBL: AB004696; BAA23309; 1; JOINED.  
 DR EMBL: AB004697; BAA23309; 1; JOINED.  
 DR EMBL: AB004698; BAA23309; 1; JOINED.  
 DR EMBL: AB004699; BAA23309; 1; JOINED.  
 DR EMBL: AB004700; BAA23309; 1; JOINED.  
 DR EMBL: AB004701; BAA23309; 1; JOINED.  
 DR EMBL: AB004702; BAA23309; 1; JOINED.  
 DR EMBL: AB004703; BAA23309; 1; JOINED.  
 DR EMBL: AB004704; BAA23309; 1; JOINED.  
 DR EMBL: AB004705; BAA23309; 1; JOINED.  
 DR EMBL: AB004706; BAA23309; 1; JOINED.  
 DR EMBL: AB004707; BAA23309; 1; JOINED.  
 DR EMBL: AB004708; BAA23309; 1; JOINED.  
 DR EMBL: AB004709; BAA23309; 1; JOINED.  
 DR EMBL: AB004710; BAA23309; 1; JOINED.  
 DR EMBL: AB004711; BAA23309; 1; JOINED.  
 DR EMBL: AB004712; BAA23309; 1; JOINED.  
 DR EMBL: AB004713; BAA23309; 1; JOINED.  
 DR EMBL: AB004714; BAA23309; 1; JOINED.  
 DR EMBL: AB004715; BAA23309; 1; JOINED.  
 DR EMBL: AB004716; BAA23309; 1; JOINED.  
 DR EMBL: AB004717; BAA23309; 1; JOINED.  
 DR EMBL: AB004718; BAA23309; 1; JOINED.  
 DR EMBL: AB004719; BAA23309; 1; JOINED.  
 DR EMBL: AB004720; BAA23309; 1; JOINED.  
 DR EMBL: AB004721; BAA23309; 1; JOINED.  
 DR EMBL: AB004722; BAA23309; 1; JOINED.  
 DR EMBL: AB004723; BAA23309; 1; JOINED.  
 DR EMBL: AB004724; BAA23309; 1; JOINED.  
 DR EMBL: AB004725; BAA23309; 1; JOINED.  
 DR EMBL: AB004726; BAA23309; 1; JOINED.  
 DR EMBL: AB004727; BAA23309; 1; JOINED.  
 DR EMBL: AB004728; BAA23309; 1; JOINED.  
 DR EMBL: AB004729; BAA23309; 1; JOINED.  
 DR EMBL: AB004730; BAA23309; 1; JOINED.  
 DR EMBL: AB004731; BAA23309; 1; JOINED.  
 DR EMBL: AB004732; BAA23309; 1; JOINED.  
 DR EMBL: AB004733; BAA23309; 1; JOINED.  
 DR EMBL: AB004734; BAA23309; 1; JOINED.  
 DR EMBL: AB004735; BAA23309; 1; JOINED.  
 DR EMBL: AB004736; BAA23309; 1; JOINED.  
 DR EMBL: AB004737; BAA23309; 1; JOINED.  
 DR EMBL: AB004738; BAA23309; 1; JOINED.  
 DR EMBL: AB004739; BAA23309; 1; JOINED.  
 DR EMBL: AB004740; BAA23309; 1; JOINED.  
 DR EMBL: AB004741; BAA23309; 1; JOINED.  
 DR EMBL: AB004742; BAA23309; 1; JOINED.  
 DR EMBL: AB004743; BAA23309; 1; JOINED.  
 DR EMBL: AB004744; BAA23309; 1; JOINED.  
 DR EMBL: AB004745; BAA23309; 1; JOINED.  
 DR EMBL: AB004746; BAA23309; 1; JOINED.  
 DR EMBL: AB004747; BAA23309; 1; JOINED.  
 DR EMBL: AB004748; BAA23309; 1; JOINED.  
 DR EMBL: AB004749; BAA23309; 1; JOINED.  
 DR EMBL: AB004750; BAA23309; 1; JOINED.  
 DR EMBL: AB004751; BAA23309; 1; JOINED.  
 DR EMBL: AB004752; BAA23309; 1; JOINED.  
 DR EMBL: AB004753; BAA23309; 1; JOINED.  
 DR EMBL: AB004754; BAA23309; 1; JOINED.  
 DR EMBL: AB004755; BAA23309; 1; JOINED.  
 DR EMBL: AB004756; BAA23309; 1; JOINED.  
 DR EMBL: AB004757; BAA23309; 1; JOINED.  
 DR EMBL: AB004758; BAA23309; 1; JOINED.  
 DR EMBL: AB004759; BAA23309; 1; JOINED.  
 DR EMBL: AB004760; BAA23309; 1; JOINED.  
 DR EMBL: AB004761; BAA23309; 1; JOINED.  
 DR EMBL: AB004762; BAA23309; 1; JOINED.  
 DR EMBL: AB004763; BAA23309; 1; JOINED.  
 DR EMBL: AB004764; BAA23309; 1; JOINED.  
 DR EMBL: AB004765; BAA23309; 1; JOINED.  
 DR EMBL: AB004766; BAA23309; 1; JOINED.  
 DR EMBL: AB004767; BAA23309; 1; JOINED.  
 DR EMBL: AB004768; BAA23309; 1; JOINED.  
 DR EMBL: AB004769; BAA23309; 1; JOINED.  
 DR EMBL: AB004770; BAA23309; 1; JOINED.  
 DR EMBL: AB004771; BAA23309; 1; JOINED.  
 DR EMBL: AB004772; BAA23309; 1; JOINED.  
 DR EMBL: AB004773; BAA23309; 1; JOINED.  
 DR EMBL: AB004774; BAA23309; 1; JOINED.  
 DR EMBL: AB004775; BAA23309; 1; JOINED.  
 DR EMBL: AB004776; BAA23309; 1; JOINED.  
 DR EMBL: AB004777; BAA23309; 1; JOINED.  
 DR EMBL: AB004778; BAA23309; 1; JOINED.  
 DR EMBL: AB004779; BAA23309; 1; JOINED.  
 DR EMBL: AB004780; BAA23309; 1; JOINED.  
 DR EMBL: AB004781; BAA23309; 1; JOINED.  
 DR EMBL: AB004782; BAA23309; 1; JOINED.  
 DR EMBL: AB004783; BAA23309; 1; JOINED.  
 DR EMBL: AB004784; BAA23309; 1; JOINED.  
 DR EMBL: AB004785; BAA23309; 1; JOINED.  
 DR EMBL: AB004786; BAA23309; 1; JOINED.  
 DR EMBL: AB004787; BAA23309; 1; JOINED.  
 DR EMBL: AB004788; BAA23309; 1; JOINED.  
 DR EMBL: AB004789; BAA23309; 1; JOINED.  
 DR EMBL: AB004790; BAA23309; 1; JOINED.  
 DR EMBL: AB004791; BAA23309; 1; JOINED.  
 DR EMBL: AB004792; BAA23309; 1; JOINED.  
 DR EMBL: AB004793; BAA23309; 1; JOINED.  
 DR EMBL: AB004794; BAA23309; 1; JOINED.  
 DR EMBL: AB004795; BAA23309; 1; JOINED.  
 DR EMBL: AB004796; BAA23309; 1; JOINED.  
 DR EMBL: AB004797; BAA23309; 1; JOINED.  
 DR EMBL: AB004798; BAA23309; 1; JOINED.  
 DR EMBL: AB004799; BAA23309; 1; JOINED.  
 DR EMBL: AB004800; BAA23309; 1; JOINED.  
 DR EMBL: AB004801; BAA23309; 1; JOINED.  
 DR EMBL: AB004802; BAA23309; 1; JOINED.  
 DR EMBL: AB004803; BAA23309; 1; JOINED.  
 DR EMBL: AB004804; BAA23309; 1; JOINED.  
 DR EMBL: AB004805; BAA23309; 1; JOINED.  
 DR EMBL: AB004806; BAA23309; 1; JOINED.  
 DR EMBL: AB004807; BAA23309; 1; JOINED.  
 DR EMBL: AB004808; BAA23309; 1; JOINED.  
 DR EMBL: AB004809; BAA23309; 1; JOINED.  
 DR EMBL: AB004810; BAA23309; 1; JOINED.  
 DR EMBL: AB004811; BAA23309; 1; JOINED.  
 DR EMBL: AB004812; BAA23309; 1; JOINED.  
 DR EMBL: AB004813; BAA23309; 1; JOINED.  
 DR EMBL: AB004814; BAA23309; 1; JOINED.  
 DR EMBL: AB004815; BAA23309; 1; JOINED.  
 DR EMBL: AB004816; BAA23309; 1; JOINED.  
 DR EMBL: AB004817; BAA23309; 1; JOINED.  
 DR EMBL: AB004818; BAA23309; 1; JOINED.  
 DR EMBL: AB004819; BAA23309; 1; JOINED.  
 DR EMBL: AB004820; BAA23309; 1; JOINED.  
 DR EMBL: AB004821; BAA23309; 1; JOINED.  
 DR EMBL: AB004822; BAA23309; 1; JOINED.  
 DR EMBL: AB004823; BAA23309; 1; JOINED.  
 DR EMBL: AB004824; BAA23309; 1; JOINED.  
 DR EMBL: AB004825; BAA23309; 1; JOINED.  
 DR EMBL: AB004826; BAA23309; 1; JOINED.  
 DR EMBL: AB004827; BAA23309; 1; JOINED.  
 DR EMBL: AB004828; BAA23309; 1; JOINED.  
 DR EMBL: AB004829; BAA23309; 1; JOINED.  
 DR EMBL: AB004830; BAA23309; 1; JOINED.  
 DR EMBL: AB004831; BAA23309; 1; JOINED.  
 DR EMBL: AB004832; BAA23309; 1; JOINED.  
 DR EMBL: AB004833; BAA23309; 1; JOINED.  
 DR EMBL: AB004834; BAA23309; 1; JOINED.  
 DR EMBL: AB004835; BAA23309; 1; JOINED.  
 DR EMBL: AB004836; BAA23309; 1; JOINED.  
 DR EMBL: AB004837; BAA23309; 1; JOINED.  
 DR EMBL: AB004838; BAA23309; 1; JOINED.  
 DR EMBL: AB004839; BAA23309; 1; JOINED.  
 DR EMBL: AB004840; BAA23309; 1; JOINED.  
 DR EMBL: AB004841; BAA23309; 1; JOINED.  
 DR EMBL: AB004842; BAA23309; 1; JOINED.  
 DR EMBL: AB004843; BAA23309; 1; JOINED.  
 DR EMBL: AB004844; BAA23309; 1; JOINED.  
 DR EMBL: AB004845; BAA23309; 1; JOINED.  
 DR EMBL: AB004846; BAA23309; 1; JOINED.  
 DR EMBL: AB004847; BAA23309; 1; JOINED.  
 DR EMBL: AB004848; BAA23309; 1; JOINED.  
 DR EMBL: AB004849; BAA23309; 1; JOINED.  
 DR EMBL: AB004850; BAA23309; 1; JOINED.  
 DR EMBL: AB004851; BAA23309; 1; JOINED.  
 DR EMBL: AB004852; BAA23309; 1; JOINED.  
 DR EMBL: AB004853; BAA23309; 1; JOINED.  
 DR EMBL: AB004854; BAA23309; 1; JOINED.  
 DR EMBL: AB004855; BAA23309; 1; JOINED.  
 DR EMBL: AB004856; BAA23309; 1; JOINED.  
 DR EMBL: AB004857; BAA23309; 1; JOINED.  
 DR EMBL: AB004858; BAA23309; 1; JOINED.  
 DR EMBL: AB004859; BAA23309; 1; JOINED.  
 DR EMBL: AB004860; BAA23309; 1; JOINED.  
 DR EMBL: AB004861; BAA23309; 1; JOINED.  
 DR EMBL: AB004862; BAA23309; 1; JOINED.  
 DR EMBL: AB004863; BAA23309; 1; JOINED.  
 DR EMBL: AB004864; BAA23309; 1; JOINED.  
 DR EMBL: AB004865; BAA23309; 1; JOINED.  
 DR EMBL: AB004866; BAA23309; 1; JOINED.  
 DR EMBL: AB004867; BAA23309; 1; JOINED.  
 DR EMBL: AB004868; BAA23309; 1; JOINED.  
 DR EMBL: AB004869; BAA23309; 1; JOINED.  
 DR EMBL: AB004870; BAA23309; 1; JOINED.  
 DR EMBL: AB004871; BAA23309; 1; JOINED.  
 DR EMBL: AB004872; BAA23309; 1; JOINED.  
 DR EMBL: AB004873; BAA23309; 1; JOINED.  
 DR EMBL: AB004874; BAA23309; 1; JOINED.  
 DR EMBL: AB004875; BAA23309; 1; JOINED.  
 DR EMBL: AB004876; BAA23309; 1; JOINED.  
 DR EMBL: AB004877; BAA23309; 1; JOINED.  
 DR EMBL: AB004878; BAA23309; 1; JOINED.  
 DR EMBL: AB004879; BAA23309; 1; JOINED.  
 DR EMBL: AB004880; BAA23309; 1; JOINED.  
 DR EMBL: AB004881; BAA23309; 1; JOINED.  
 DR EMBL: AB004882; BAA23309; 1; JOINED.  
 DR EMBL: AB004883; BAA23309; 1; JOINED.  
 DR EMBL: AB004884; BAA23309; 1; JOINED.  
 DR EMBL: AB004885; BAA23309; 1; JOINED.  
 DR EMBL: AB004886; BAA23309; 1; JOINED.  
 DR EMBL: AB004887; BAA23309; 1; JOINED.  
 DR EMBL: AB004888; BAA23309; 1; JOINED.  
 DR EMBL: AB004889; BAA23309; 1; JOINED.  
 DR EMBL: AB004890; BAA23309; 1; JOINED.  
 DR EMBL: AB004891; BAA23309; 1; JOINED.  
 DR EMBL: AB004892; BAA23309; 1; JOINED.  
 DR EMBL: AB004893; BAA23309; 1; JOINED.  
 DR EMBL: AB004894; BAA23309; 1; JOINED.  
 DR EMBL: AB004895; BAA23309; 1; JOINED.  
 DR EMBL: AB004896; BAA23309; 1; JOINED.  
 DR EMBL: AB004897; BAA23309; 1; JOINED.  
 DR EMBL: AB004898; BAA23309; 1; JOINED.  
 DR EMBL: AB004899; BAA23309; 1; JOINED.  
 DR EMBL: AB004900; BAA23309; 1; JOINED.  
 DR EMBL: AB004901; BAA23309; 1; JOINED.  
 DR EMBL: AB004902; BAA23309; 1; JOINED.  
 DR EMBL: AB004903; BAA23309; 1; JOINED.  
 DR EMBL: AB004904; BAA23309; 1; JOINED.  
 DR EMBL: AB004905; BAA23309; 1; JOINED.  
 DR EMBL: AB004906; BAA23309; 1; JOINED.  
 DR EMBL: AB004907; BAA23309; 1; JOINED.  
 DR EMBL: AB004908; BAA23309; 1; JOINED.  
 DR EMBL: AB004909; BAA23309; 1; JOINED.  
 DR EMBL: AB004910; BAA23309; 1; JOINED.  
 DR EMBL: AB004911; BAA23309; 1; JOINED.  
 DR EMBL: AB004912; BAA23309; 1; JOINED.  
 DR EMBL: AB004913; BAA23309; 1; JOINED.  
 DR EMBL: AB004914; BAA23309; 1; JOINED.  
 DR EMBL: AB004915; BAA23309; 1; JOINED.  
 DR EMBL: AB004916; BAA23309; 1; JOINED.  
 DR EMBL: AB004917; BAA23309; 1; JOINED.  
 DR EMBL: AB004918; BAA23309; 1; JOINED.  
 DR EMBL: AB004919; BAA23309; 1; JOINED.  
 DR EMBL: AB004920; BAA23309; 1; JOINED.  
 DR EMBL: AB004921; BAA23309; 1; JOINED.  
 DR EMBL: AB004922; BAA23309; 1; JOINED.  
 DR EMBL: AB004923; BAA23309; 1; JOINED.  
 DR EMBL: AB004924; BAA23309; 1; JOINED.  
 DR EMBL: AB004925; BAA23309; 1; JOINED.  
 DR EMBL: AB004926; BAA23309; 1; JOINED.  
 DR EMBL: AB004927; BAA23309; 1; JOINED.  
 DR EMBL: AB004928; BAA23309; 1; JOINED.  
 DR EMBL: AB004929; BAA23309; 1; JOINED.  
 DR EMBL: AB004930; BAA23309; 1; JOINED.  
 DR EMBL: AB004931; BAA23309; 1; JOINED.  
 DR EMBL: AB004932; BAA23309; 1; JOINED.  
 DR EMBL: AB004933; BAA23309; 1; JOINED.  
 DR EMBL: AB004934; BAA23309; 1; JOINED.  
 DR EMBL: AB004935; BAA23309; 1; JOINED.  
 DR EMBL: AB004936; BAA23309; 1; JOINED.  
 DR EMBL: AB004937; BAA23309; 1; JOINED.  
 DR EMBL: AB004938; BAA23309; 1; JOINED.  
 DR EMBL: AB004939; BAA23309; 1; JOINED.  
 DR EMBL: AB004940; BAA23309; 1; JOINED.  
 DR EMBL: AB004941; BAA23309; 1; JOINED.  
 DR EMBL: AB004942; BAA23309; 1; JOINED.  
 DR EMBL: AB004943; BAA23309; 1; JOINED.  
 DR EMBL: AB004944; BAA23309; 1; JOINED.  
 DR EMBL: AB004945; BAA23309; 1; JOINED.  
 DR EMBL: AB004946; BAA23309; 1; JOINED.  
 DR EMBL: AB004947; BAA23309; 1; JOINED.  
 DR EMBL: AB004948; BAA23309; 1; JOINED.  
 DR EMBL: AB004949; BAA23309; 1; JOINED.  
 DR EMBL: AB004950; BAA23309; 1; JOINED.  
 DR EMBL: AB004951; BAA23309; 1; JOINED.  
 DR EMBL: AB004952; BAA23309; 1; JOINED.  
 DR EMBL: AB004953; BAA23309; 1; JOINED.  
 DR EMBL: AB004954; BAA23309; 1; JOINED.  
 DR EMBL: AB004955; BAA23309; 1; JOINED.  
 DR EMBL: AB004956; BAA23309; 1; JOINED.  
 DR EMBL: AB004957; BAA23309; 1; JOINED.  
 DR EMBL: AB004958; BAA23309; 1; JOINED.  
 DR EMBL: AB004959; BAA23309; 1; JOINED.  
 DR EMBL: AB004960; BAA23309; 1; JOINED.  
 DR EMBL: AB004961; BAA23309; 1; JOINED.<br

Db 457 |:||||| | ||| | GLKDKGEGGLPGFP 471

RESULT 29

CALC\_MOUSE STANDARD; PRT; 1464 AA.

ID CA13\_MOUSE STANDARD; PRT; 1464 AA.

AC P08121; 061429; Q9GRN7;

DT 01-AUG-1988 (Rel. 08, created)

DT 15-JUL-1999 (Rel. 38, last sequence update)

DT 16-OCT-2001 (Rel. 40, last annotation update)

DE Collagen alpha I(III) chain precursor.

GN COL1A1.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathii; Muridae; Murinae; Mus.

OX NCBI\_TaxID=10090;

RN [1]

RL PRT; 1464 AA.

RP SEQUENCE FROM N.A.

RC STRAIN=C7BL X DBA; TISSUE=Embryo;

RX MEDLINE=55011609; PubMed=7920795;

RA Toman D., de Crombrugge B.;

RT "The mouse type-I/II procollagen-encoding gene: genomic cloning and complete DNA sequence.";

RT Gene 147:161-168(1994).

RN [2]

RP SEQUENCE OF 1-488 FROM N.A.

RC MEDLINE=81167858; PubMed=3443309;

RX Wood L., Theriault N., Vogeli G.;

RA "Complete nucleotide sequence of the N-terminal domains of the murine alpha-1 type-I/II collagen chain.";

RT Gene 61:225-230(1987).

RL PRT; 1464 AA.

RN [3]

RP SEQUENCE OF 1-28 FROM N.A.

RC MEDLINE=5131189; PubMed=3972847;

RA Liu G., Mudryj M., de Crombrugge B.;

RT "Identification of the promoter and first exon of the mouse alpha 1 (III) collagen gene.;"

RL J. Biol. Chem. 260:3773-3777(1985).

RN [4]

RP SEQUENCE OF 810-1464 FROM N.A.

RC STRAIN=C7BL/6J; TISSUE=Embryonic head;

RX MEDLINE=3105660; PubMed=11217851;

RA Kawai J., Shinagawa A., Shihata K., Yoshino M., Itoh M., Ishii Y., Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S., Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamamoto I., Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R., Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T., Fleischmann R., Gaasterland T., Gissi C., King B., Kochiwa H., Kuehl P., Lewis M., Matsuo Y., Nikaido T., Pesole G., Quackenbush J., Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T., Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Baish G., Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F., Brownstein M.J., Bult C., Fletcher C., Fujita M., Garibaldi M., Gustincich S., Hill D., Hofmann M., Hume D.A., Kameya M., Lee N.H., Lyons P., Marchionni L., Mashima J., Mazzarella J., Mombaerts P., Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N., Sasaki H., Sato K., Schoenbach C., Sepe T., Shihata Y., Storch K.-F., Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L., Wynnshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsuki S., Hayashizaki Y.;

RA "Functional annotation of a full-length mouse cDNA collection.";

RL Nature 409:665-690(2001).

RN [5]

RP SEQUENCE OF 1442-1464 FROM N.A.

RC STRAIN=C7BL;

RX MEDLINE=61274355; PubMed=2054384;

RA Mettaerranta M., Toman D., de Crombrugge B., Vuorio E.;

RT "Specific hybridization probes for mouse type I, II, III and IX collagen mRNAs.";

RL Biophys. Acta 1089:241-243(1991).

CC -!- FUNCTION: COLLAGEN TYPE III OCCURS IN MOST SOFT CONNECTIVE TISSUES ALONG WITH TYPE I COLLAGEN.

CC -!- SUBUNIT: TRIMERS OF IDENTICAL ALPHA 1(III) CHAINS. THE CHAINS ARE LINKED TO EACH OTHER BY INTERCHAIN DISULFIDE BONDS. TRIMERS ARE ALSO CROSS-LINKED VIA HYDROXYLINES.

CC -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPETIDE REPEATING UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.

CC O-LINKED GLYCAN CONSISTS OF GLC-GAL DISACCHARIDE (BY SIMILARITY).

CC -!- SIMILARITY: CONTAINS 1 WFCC DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See [http://www.isb-sib.ch/announce/or\\_send\\_an\\_email\\_to\\_licensee@isb-sib.ch](http://www.isb-sib.ch/announce/or_send_an_email_to_licensee@isb-sib.ch)).

CC -----

DR EMBL; X52046; CAA3279\_1; -;

DR EMBL; M18933; AAA31338\_1; -;

DR EMBL; K03037; -; NOT\_ANNOTATED\_CDS.

DR EMBL; AK01948; BAB31724\_1; -;

DR EMBL; X51983; CAA41048\_1; -;

DR PIR; A22287; A22287.

DR PIR; A22353; A22353.

DR PIR; S16373; S16373.

DR MGD; MGI:88453; Col3a1.

DR InterPro; IPR00087; Collagen.

DR InterPro; IPR00885; Fib.collagen\_C.

DR Pfam; PF01410; COLF1; 1.

DR Pfam; PF01391; Collagen; 17.

DR Prodom; PD002078; Fib.collagen\_C; 1.

DR SMART; SM00214; VWC; 1.

DR PROSITE; PS01208; VWFC; 1.

KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation; Glycoprotein; Collagen; Signal.

FT SIGNAL 1 23 BY SIMILARITY.

FT PROPER 24 154 AMINO-TERMINAL PROPEPTIDE.

FT CHAIN 155 1203 COLLAGEN ALPHA 1(III) CHAIN.

FT PROPEP 1204 1464 CARBOXYL-TERMINAL PROPEPTIDE.

FT DOMAIN 31 90 VWFCC.

FT DOMAIN 155 169 NONHELICAL REGION (N-TERMINAL).

FT DOMAIN 170 1195 TRIPLE-HELICAL REGION.

FT DOMAIN 196 1464 NONHELICAL REGION (C-TERMINAL).

FT CARBOHYD 262 262 O-LINKED GAL. . . (BY SIMILARITY).

FT MOD\_RES 262 262 HYDROXYLATION (BY SIMILARITY).

FT MOD\_RES 283 283 HYDROXYLATION (BY SIMILARITY).

FT MOD\_RES 859 859 HYDROXYLATION (BY SIMILARITY).

FT MOD\_RES 976 976 HYDROXYLATION (BY SIMILARITY).

FT MOD\_RES 1093 1093 HYDROXYLATION (BY SIMILARITY).

FT MOD\_RES 1105 1105 HYDROXYLATION (BY SIMILARITY).

FT DISULFID 1195 1195 INTERCHAIN (BY SIMILARITY).

FT DISULFID 1196 1196 INTERCHAIN (BY SIMILARITY).

SQ SEQUENCE 1464 AA; 138944 MW; 2104EC27A886090B CRC64;

Query Match 60.9%; Score 56; DB 1; Length 1464;

Best Local Similarity 69.2%; Pred. No. 1.1;

Matches 9; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 GVKGDKGNWGPG 13

1111:1:1:1:1

Db 857 GVKGDRGSPGGPG 869

RESULT 30

MSRE\_MOUSE STANDARD; PRT; 458 AA.

ID P30204; Q9ZG6;

AC P30204;

DT 01-APR-1993 (Rel. 25, created)

DT 01-OCT-1996 (Rel. 34, last sequence update)

DT 01-MAR-2002 (Rel. 41, last annotation update)

DE Macrophage scavenger receptor types I and II (Macrophage acetylated LDL receptor I and II) (Scavenger receptor type A) (SR-A).

GN MSRI OR SCVR.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TAXID=10090;  
 RN [1] SEQUENCE FROM N.A. (ISOFORMS I AND II).  
 RP MEDLINE=93359822; Published=8394868;  
 RA Ashkenas J., Penman M., Vasile E., Acton S., Freeman M.W.,  
 Krieger M.;  
 RT "Structures and high and low affinity ligand binding properties of  
 murine type I and type II macrophage scavenger receptors.";  
 RL J. Lipid Res. 34: 983-1000(1993).  
 RN [2] SEQUENCE FROM N.A. (ISOFORM III).  
 RP MEDLINE=93131972; Published=8380589;  
 RA DOI T., Wada Y., Kodama T., Higashii K.I., Kurihara Y.,  
 Miyazaki T., Nakamura H., Uesugi S., Imanishi T., Kawabe Y.,  
 Itakura H., Yazaki Y., Matsumoto A.;  
 RT "Charged collagen structure mediates the recognition of negatively  
 charged macromolecules by macrophage scavenger receptors.";  
 RL J. Biol. Chem. 268:2126-2133(1993).  
 RN [3] SEQUENCE FROM N.A. (ISOFORM I).  
 RP STRAIN=C57BL/6;  
 RA Rateri D.L., Whitman S.C., Block A.E., Daugherty A.;  
 RT "Identification of a functional domain in class A scavenger receptors  
 that mediates metabolism of AcLDL";  
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
 RN [4] SEQUENCE OF 349-458 FROM N.A.  
 RP MEDLINE=91062370; Published=1978939;  
 RA Freeman M., Ashkenas J., Kingsley D.M., Copeland N.G.,  
 Jenkins N.A., Krieger M.;  
 RT "An ancient, highly conserved family of cysteine-rich protein domains  
 revealed by cloning type I and type II murine macrophage scavenger  
 receptors.;"  
 RT Proc. Natl. Acad. Sci. U.S.A. 87:8810-8814(1990).  
 RN [5] SEQUENCE OF 1-4 FROM N.A.  
 RP MEDLINE=9535388; Published=76666008;  
 RA Affring R.P., Freeman M.W.;  
 RT "structure of the murine macrophage scavenger receptor gene and  
 evaluation of sequences that regulate expression in the macrophage  
 cell line, P388D;"  
 RL J. Lipid Res. 36:1305-1314(1995).  
 CC -I- FUNCTION: MEMBRANE GLYCOPROTEINS IMPLICATED IN THE PATHOLOGIC  
 DEPOSITION OF CHOLESTEROL IN ARTERIAL WALLS DURING ATHEROGENESIS.  
 TWO TYPES OF RECEPTOR SUBUNITS EXIST. THESE RECEPTORS MEDIATE THE  
 ENDOCYTOSIS OF A DIVERSE GROUP OF MACROMOLECULES, INCLUDING  
 MODIFIED LOW DENSITY LIPOPROTEINS (LDL).  
 -I- SUBUNIT: HOMOTIMER.  
 CC -I- SUBCELLULAR LOCATION: Type II membrane protein.  
 -I- ALTERNATIVE PRODUCTS: 2 ISOFORMS; I (SHOWN HERE) AND II; ARE  
 CC PRODUCED BY ALTERNATIVE SPLICING.  
 CC -I- SIMILARITY: CONTAINS 1 SRCR DOMAIN.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL; L04275; AAA39748.1; ALT\_INIT.  
 DR EMBL; D13382; BAA02650.1; .  
 DR EMBL; AF203181; AF14001.1; .  
 DR EMBL; M59445; AAA37464.1; .  
 DR EMBL; M5945; AA337465.1; .  
 DR EMBL; U13873; AAC13774.1; .  
 DR PIR; A38260; A38260.

DR MGD; MGI\_98257; Msrl.  
 DR InterPro; IPR000087; Collagen.  
 DR InterPro; IPR01190; SRCR.  
 DR Pfam; PF01391; Collagen\_1.  
 DR Pfam; PF00530; SRCR; I.  
 DR PRINTS; PR00258; SPERACTRCPR.  
 DR SMART; SM00202; SR; I.  
 DR PROSITE; PS00420; SRCR\_1; I.  
 DR PROSITE; PS00427; SRCR\_2; I.  
 KW Transmembrane; Glycoprotein; Receptor; Signal-anchor; Alternative splicing; CYTROPLASMIC (POTENTIAL).  
 KW TRANSMEM 56 78 55 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)  
 FT DOMAIN 357 457 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)  
 FT DOMAIN 79 458 (POTENTIAL).  
 FT DOMAIN 79 114 EXTRACELLULAR (POTENTIAL).  
 FT DOMAIN 209 259 SPACER (PROBABLE).  
 FT DOMAIN 209 259 COILED COIL (POTENTIAL).  
 FT DOMAIN 277 350 COLLAGEN-LIKE.  
 FT CARBOHYD 357 457 SRCR.  
 FT CARBOHYD 94 94 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 107 107 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 147 147 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 188 188 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 253 253 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 271 271 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT VARSPLC 352 354 TPL -> RSV (IN ISOFORM II).  
 FT VARSPLC 355 458 MISSING (IN ISOFORM II).  
 FT CONFLICT 47 47 V -> L (IN REF. 3).  
 FT CONFLICT 110 110 N -> K (IN REF. 3).  
 FT CONFLICT 120 120 A -> E (IN REF. 3).  
 FT CONFLICT 130 130 O -> E (IN REF. 3).  
 FT CONFLICT 168 168 L -> S (IN REF. 3).  
 FT CONFLICT 202 202 H -> N (IN REF. 3).  
 SQ SEQUENCE 458 AA; 50130 MW; 6D4C01C27BEE50B CRG64;

Query	Match	Best Local Similarity	Score	DB	Length
Qy	1 GVKGD-----KGPGWPGAP 15	50.0%	55.5;	DB	458
Db	307 GVKGDGQIGFPGGRGNFGNGAP 330				

Search completed: November 1, 2002, 12:52:53  
 Job time : 9 secs

**THIS PAGE BLANK (USP TO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

## OM protein - protein search, using sw model.

Run on: November 1, 2002, 12:19:21 ; Search time 21.5 Seconds

(without alignments) (120.694 Million cell updates/sec)

## Title: US-09-529-691A-1

Perfect score: 92

Sequence: 1 GVKGDKGNPGWPGAP 15

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL\_19;\*

1: sp\_archea;\*  
2: sp\_bacteria;\*  
3: sp\_fungi;\*  
4: sp\_human;\*  
5: sp\_invertebrate;\*  
6: sp\_mammal;\*  
7: sp\_mhc;\*  
8: sp\_organelle;\*  
9: sp\_phage;\*  
10: sp\_plant;\*  
11: sp\_rabbit;\*  
12: sp\_virus;\*  
13: sp\_vertebrate;\*  
14: sp\_unclassified;\*  
15: sp\_rvirus;\*  
16: sp\_bacteriaph;\*  
17: sp\_archeap;\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	92	100.0	1621	4 Q9H4R9
2	69	75.0	806	4 Q9NUB7
3	66	71.7	404	2 Q9F8E7
4	64	69.6	673	4 Q1A052
5	64	69.6	1691	11 Q9PSQ2
6	63	68.5	174	13 Q90585
7	63	68.5	979	13 Q919K3
8	62	67.4	290	5 Q26054
9	62	67.4	412	9 Q9FYX3
10	62	67.4	1414	5 Q26634
11	62	67.4	1870	4 Q9BQT2
12	62	67.4	1802	5 Q17163
13	61	66.3	224	5 Q9VYV2
14	61	66.3	447	4 Q16593
15	61	66.3	957	4 Q9HOV3
16	66.3	957	4 Q96P44	

Database : SPTREMBL\_19;\*

1: sp\_archea;\*  
2: sp\_bacteria;\*  
3: sp\_fungi;\*  
4: sp\_human;\*  
5: sp\_invertebrate;\*  
6: sp\_mammal;\*  
7: sp\_mhc;\*  
8: sp\_organelle;\*  
9: sp\_phage;\*  
10: sp\_plant;\*  
11: sp\_rabbit;\*  
12: sp\_virus;\*  
13: sp\_vertebrate;\*  
14: sp\_unclassified;\*  
15: sp\_rvirus;\*  
16: sp\_bacteriaph;\*  
17: sp\_archeap;\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## ALIGNMENTS

RESULT 1									
Q9H4R9	PRELIMINARY;	PRT;	1621 AA.						
ID Q9H4R9	Q9H4R9; 01-MAR-2001 (TREMBLrel. 16, Created)								
AC Q9H4R9;	DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)								
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)	DE BA472K17.2 (COLLAGEN TYPE IV ALPHA 1) (FRAGMENT).								
GN COLA1.	OS Homo sapiens (Human).								
RA RPT; C4; 2.	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.								
DR NEBL_TAXID=9606;	RN [1]								
RP SEQUENCE FROM N.A.	RA RPT; C4; 2.								
RA RPT; C4; 2.	RL Submitted (OCT-2000) to the EMBL/GenBank/DDBJ databases.								
DR EMBL; AL390755; CACI3153.1; -	DR InterPro; IPR00087; C4.								
DR InterPro; IPR00087; Collagen.	DR InterPro; IPR00087; Collagen.								
DR PFam; PF01413; C4; 2.	DR PRODOM; PDD03923; C4; 2.								
DR PRODOM; PDD03923; C4; 2.	DR SMART; SM00111; C4; 2.								
KW Collagen.	FT NON_TER								
FT NON_TER	SQ 1621 AA; 155705 MW; 73P6FF901CD0EDBA2 CRC64;								
SQ 1621 AA;	SQ NON_TER 1 1621 AA;								
Query Match Best Local Similarity Matches	100.0%; Score 92; DB 4; Length 1621; 100.0%; Pred. No. 1.5e-05; 0; Mismatches 0; Indels 0; Gaps 0;								
QY 1 Q9H4R9	Db 1215 GVKGDKGNPGWPGAP 1229								
RESULT 2	ID Q9NUB7 PRELIMINARY; PRT; 886 AA.								

AC Q9NQB7; DT 01-OCT-2000 (TREMBIREL. 15, Created)  
 AC Q14052; DT 01-OCT-2000 (TREMBIREL. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMBIREL. 19, Last annotation update)  
 DE DA2A23\_1 (COLLAGEN, TYPE IV, ALPHA 5 (ALPORT SYNDROME))  
 DE (FRAGMENT).  
 GN COLA5.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606; [1]  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Cobley V.;  
 RL Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AL035425; CAM90289.1; -.  
 DR InterPro; IPR001442; C4; Collagen.  
 DR InterPro; IPR000087; Collagen.  
 DR Pfam; PF01413; C4; 2.  
 DR Pfam; PF01391; Collagen; 10.  
 DR ProDom; PD000323; C4; 2.  
 DR SMART; SM00111; C4; 2.  
 KW Collagen.  
 FT NON\_TER 1 1  
 SQ SEQUENCE 886 AA; 854/9 MW; 8C06B9FC9A9A6569 CRC64;  
 Query Match 75.0%; Score 69; DB 4; Length 886;  
 Best Local Similarity 73.3%; Pred. No. 0.027; 11; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
 Db 480 GIKGEGKGNPGQPGIP 494

---

RESULT 3  
 Q9FF687  
 ID Q9F687  
 AC Q9F687;  
 DT 01-MAR-2001 (TREMBIREL. 16, Created)  
 DT 01-MAR-2001 (TREMBIREL. 16, Last sequence update)  
 DT 01-JUN-2001 (TREMBIREL. 17, Last annotation update)  
 DE COLLAGEN-LIKE SURFACE PROTEIN (FRAGMENT).  
 OS Streptococcus pyogenes.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 OC Streptococcus.  
 OX NCBI\_TaxID=1314;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A34;  
 RX MEDLINE=20490571; PubMed=11035747;  
 RA Rasmussen M., Eden A., Bjorck L.;  
 RT "SclC, a novel collagen-like surface protein of streptococcus  
 pyogenes.,"  
 RL Infect. Immun. 68:6337-6377(2000).  
 DR EMBL; AF296336; AAC30215.1; -.  
 DR InterPro; IPR000087; Collagen.  
 DR Pfam; PF01391; Collagen; 3.  
 FT NON\_TER 1  
 FT NON\_TER 404  
 SQ SEQUENCE 404 AA; 41612 MW; 45304C89E3B41A78 CRC64;

Query Match 71.7%; Score 66; DB 2; Length 404;  
 Best Local Similarity 73.3%; Pred. No. 0.034; 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

---

RESULT 4  
 QY 1 GVKGDKGNPGWPGAP 15  
 QY 1 :|||||1:|||||1:|||||1:|||||1:  
 Db 329 GPKGPKGDPGFPGPAP 343

---

RESULT 5  
 Q9ESQ2  
 ID Q9ESQ2  
 AC Q9ESQ2;  
 DT 01-MAR-2001 (TREMBIREL. 16, Created)  
 DT 01-MAR-2001 (TREMBIREL. 16, Last sequence update)  
 DT 01-JUN-2001 (TREMBIREL. 17, Last annotation update)  
 DE TYPE IV COLLAGEN ALPHA 5 CHAIN.  
 GN COLA5.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090; [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20536494; PubMed=110965041;  
 RA Saito K., Naito T., Seki T., Oohashi T., Kimura E., Momota R.,  
 RA Kishino Y., Saito Y., Yoshioka H., Ninomiya Y.;  
 RT "Differential Expression of Mouse a5(IV) and a6(IV) Collagen Genes in  
 RT Epithelial Basement Membranes.,"  
 RL J. Biochem. 128:427-434(2000).  
 DR EMBL; AB041350; BAB13673.1; -.  
 DR InterPro; IPR001442; C4; Collagen.  
 DR InterPro; IPR000087; Collagen.  
 DR Pfam; PF01413; C4; 2.  
 DR ProDom; PD000323; C4; 2.  
 DR SMART; SM00111; C4; 2.  
 KW Collagen.  
 SQ SEQUENCE 1691 AA; 161823 MW; 81340DF1792208FA CRC64;

Query Match 69.6%; Score 64; DB 11; Length 1691;  
 Best Local Similarity 66.7%; Pred. No. 0.3; 10; Conservative 3; Mismatches 2; Indels 0; Gaps 0;  
 Matches 10; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GVKGDKNPGWPGAP 15  
ID ||||:|||||  
AC 090585 PRELIMINARY; PRT; 174 AA.  
DB 1285 GIKGEGNPGPPGQP 1299

## RESULT 6

Q90585 PRELIMINARY; PRT; 174 AA.  
ID 090585 ID 026054 PRELIMINARY; PRT; 290 AA.  
AC 090585; ID 026054; PRELIMINARY; PRT; 290 AA.  
DT 01-NOV-1996 (TREMBrel. 01, Created)  
DT 01-NOV-1996 (TREMBrel. 01, Last sequence update)  
DT 01-NOV-1996 (TREMBrel. 19, Last annotation update)  
DE CARTILAGE ALPHA-1(IX) COLLAGEN PROTEOGLYCAN (FRAGMENT).  
OS Gallus gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
OC Gallus.  
OC NCBI\_TaxID:9031;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90052114; PubMed=2584206;  
RA Nishimura I., Muragaki Y., Olsen B.R.;  
RT "Issue-specific forms of type IX collagen-proteoglycan arise from the  
use of two widely separated promoters.";  
RL J. Biol. Chem. 264:20033-20041(1989).  
DR EMBL: M28660; AA:A48705; 1. -.  
DR InterPro: IPR000087; Collagen.  
DR Pfam: PF01391; Collagen; 2.  
KW Collagen.  
FT NON\_TER 174 174  
SQ SEQUENCE 174 AA; 16240 MW; C0E64A0748A224FO CRC64;

Query Match Score 63; DB 13; Length 174;  
Best Local Similarity 66.7%; Pred. No. 0.041; Mismatches 3; Indels 0; Gaps 0;  
Matches 10; Conservative 3;

Qy 1 GVKGDKNPGWPGAP 15  
ID ||||:|||||  
DB 46 GIDGDKGSPGPSP 60

## RESULT 7

Q9I9K3 PRELIMINARY; PRT; 979 AA.  
ID Q9I9K3 ID 026054 PRELIMINARY; PRT; 290 AA.  
AC 09I9K3; ID 026054; PRELIMINARY; PRT; 290 AA.  
DT 01-OCT-2000 (TREMBrel. 15, Created)  
DT 01-OCT-2000 (TREMBrel. 15, Last sequence update)  
DT 01-JUN-2001 (TREMBrel. 17, Last annotation update)  
DE COLLAGEN IV A1 CHAIN (FRAGMENT).  
OS Gallus gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
OC Gallus.  
OC NCBI\_TaxID:9031;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Halfter W.M., Dong S.;  
RT "Composition, synthesis and assembly of the embryonic chick retinal  
basal lamina.";  
RL Dev. Biol. 0:0-0(2000);  
DR EMBL: AF239838; AA:F4681; 1. -.  
DR InterPro: IPR001442; C4.  
DR Pfam: PF01413; C4; 2.  
DR Pfam: PF01391; Collagen; 12.  
DR ProDom: PD003923; C4; 2.  
DR SMART: SM00111; C4; 2.  
FT NON\_TER 1 1  
SQ SEQUENCE 979 AA; 95020 MW; 5B1017D911ED4299 CRC64;

Query Match Score 63; DB 13; Length 979;  
Best Local Similarity 60.0%; Pred. No. 0.24; Mismatches 3; Indels 0; Gaps 0;  
Matches 9; Conservative 3;

## RESULT 8

Q9FZY3 PRELIMINARY; PRT; 412 AA.  
ID Q9FZY3 ID 026054 PRELIMINARY; PRT; 290 AA.  
AC 09FZY3; ID 026054; PRELIMINARY; PRT; 290 AA.  
DT 01-MAR-2001 (TREMBrel. 16, Created)  
DT 01-MAR-2001 (TREMBrel. 16, Last sequence update)  
DT 01-DEC-2001 (TREMBrel. 19, Last annotation update)  
DE ORF62.  
OS bacteriophage phi ETA.  
OC Viruses.  
OC NCBI\_TaxID:106284;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Sugai M., Yamaguchi T., Hayashi T., Nakasone K., Takami H.;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=2056787; PubMed=11115106;  
RA Yamaguchi T., Hayashi T., Takami H., Nakasone K., Ohnishi M.,  
RA Nakayama K., Yamada S., Komatsuwa H., Sugai M.,  
RT "Phage conversion of exfoliative toxin A production in Staphylococcus  
aureus";  
RL Mol. Microbiol. 38:594-705(2000).  
DR EMBL: AP001553; BAB97648; 1. -.  
DR InterPro: IPR00087; Collagen.  
DR ProDom: PD003923; C4; 2.  
DR SMART: SM00111; C4; 2.  
FT NON\_TER 1 1  
SQ SEQUENCE 412 AA; 46013 MW; 3E87FCBAB223088 CRC64;

Query Match Score 62; DB 9; Length 412;  
Best Local Similarity 78.6%; Pred. No. 0.14; Mismatches 3; Indels 0; Gaps 0;  
Matches 11; Conservative 0;

Qy 1 GVKGDKNPGWPGAP 14

Db 178 GAKGDKGEFGQPGA 191  
**RESULT 10**  
 ID 026534 PRELIMINARY; PRT; 1414 AA.  
 AC Q26534;  
 DT 01-NOV-1996 (TREMBREL. 01, Created)  
 DT 01-NOV-1996 (TREMBREL. 01, Last sequence update)  
 DT 01-DEC-2001 (TREMBREL. 19, Last annotation update)  
 DE ALPHA-1 COLLAGEN  
 GN COL1ALPHA.  
 OS Strongylocentrotus purpuratus (Purple sea urchin).  
 OC Eukaryota; Metazoa; Echinodermata; Eleutheriozoa; Echinozoa;  
 OC Echinoidea; Euechinoidea; Echinacea; Echinida; Strongylocentrotidae;  
 OC Strongylocentrotus.  
 NCBI\_TaxID=7658;  
 [1]  
 RP SEQUENCE FROM N.A.  
 MEDLINE=2348411; PubMed=1639795;  
 RA Exposito J.-Y.; D'Alessio M.; Solursh M.; Ramirez F.;  
 RT "Sea urchin collagen evolutionarily homologous to vertebrate pro-  
 alpha-2(1) collagen.";  
 RL J. Biol. Chem. 267:15559-15562(1992).  
 EMBL; M92040; AAA0035.1; .  
 DR InterPro; IPR000081; Collagen.  
 DR InterPro; IPR000885; Fib\_collagen\_C.  
 PFam; PF01410; COLF1; 1.  
 DR Pfam; PF01391; Collagen; 18.  
 DR Pfam; PF002078; Fib\_collagen\_C; 1.  
 DR SMART; SM00038; COLFI; 1.  
 KW Collagen. 1414 AA; 133025 MW; 94D9CBA71A9FD73D CRC64;  
 Query Match 67.4%; Score 62; DB 5; Length 1414.  
 Best Local Similarity 78.6%; Pred. No. 0.51;  
 Matches 11; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 Qy 1 GVKDQKGNPGWPGA 14  
 | |||:|||||||  
 Db 393 GSKEGDQGNPGQPGA 406  
 OS Homo sapiens (Human).  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21064696; PubMed=11134255;  
 RA Heidet L.; Arredondo C.; Forestier L.; Cohen-Solal L.; Mollet G.,  
 Gutierrez B.; Stavrou C.; Gubler M.C.; Antignac C.;  
 RT "Structure of the human type IV collagen gene COL4A3 and mutations in  
 autosomal Airport syndrome.";  
 RL J. Am. Soc. Nephrol. 12:97-106(2001).  
 EMBL; AJ288487; CAC36101.1; .  
 DR EMBL; AJ288488; CAC36101.1; JOINED.  
 DR EMBL; AJ288489; CAC36101.1; JOINED.  
 DR EMBL; AJ288490; CAC36101.1; JOINED.  
 DR EMBL; AJ288491; CAC36101.1; JOINED.  
 DR EMBL; AJ288492; CAC36101.1; JOINED.  
 DR EMBL; AJ288493; CAC36101.1; JOINED.  
 DR EMBL; AJ288494; CAC36101.1; JOINED.

**RESULT 11**  
 ID Q9BQI2 PRELIMINARY; PRT; 1670 AA.  
 AC Q9BQI2;  
 DT 01-JUN-2001 (TREMBREL. 17, Created)  
 DT 01-JUN-2001 (TREMBREL. 17, Last sequence update)  
 DT 01-OCT-2001 (TREMBREL. 18, Last annotation update)  
 DE ALPHA3 TYPE IV COLLAGEN.  
 COLA3  
 OS Homo sapiens (Human).  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21064696; PubMed=11134255;  
 RA Heidet L.; Arredondo C.; Forestier L.; Cohen-Solal L.; Mollet G.,  
 Gutierrez B.; Stavrou C.; Gubler M.C.; Antignac C.;  
 RT "Structure of the human type IV collagen gene COL4A3 and mutations in  
 autosomal Airport syndrome.";  
 RL J. Am. Soc. Nephrol. 12:97-106(2001).  
 EMBL; AJ288487; CAC36101.1; .  
 DR EMBL; AJ288488; CAC36101.1; JOINED.  
 DR EMBL; AJ288489; CAC36101.1; JOINED.  
 DR EMBL; AJ288490; CAC36101.1; JOINED.  
 DR EMBL; AJ288491; CAC36101.1; JOINED.  
 DR EMBL; AJ288492; CAC36101.1; JOINED.  
 DR EMBL; AJ288493; CAC36101.1; JOINED.  
 DR EMBL; AJ288494; CAC36101.1; JOINED.

**RESULT 12**  
 ID Q17163 PRELIMINARY; PRT; 1802 AA.  
 AC Q17163;  
 DT 01-NOV-1996 (TREMBREL. 01, Created)  
 DT 01-NOV-1996 (TREMBREL. 01, Last sequence update)  
 DT 01-DEC-2001 (TREMBREL. 19, Last annotation update)

DE A2 (IV) BASEMENT MEMBRANE COLLAGEN.	RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Mosherfi A.,
OS Brugia malayi.	RA Mount S.M., Moy M., Murphy B., Murphy L., Muzyk D.M., Nelson D.L.,
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;	RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
ONCIL_TAXID=6279;	RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RN [1]	RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RP SEQUENCE FROM N.A.	RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RC STRAIN=APERIODIC;	RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RX MEDLINE=9534846; PubMed=7637709;	RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Culaghi V.R., Rajan T.V.,	RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
RT "The structural organization of an alpha 2 (type IV) basement membrane	RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
collagen gene from the filarial nematode <i>Brugia malayi</i> .";	RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao O., Zheng L.,
Mol. Biochem. Parasitol. 70:227-229(1995).	RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
EMBL; U07224; AAC6611.1; -.	RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
DR InterPro; IPR01442; C4.	RA RT "The genome sequence of <i>Drosophila melanogaster</i> .";
DR InterPro; IPR00087; Collagen.	RA Science 287:2185-2195(2000).
DR Pfam; PF01413; C4; 2.	RA EMBL; AB003517; ARF49173.1.
DR ProDom; PDO3923; C4; 2.	RA DR FlyBase; FBAn056861; CG14089.
DR SMART; SM00111; C4; 2.	RA DR InterPro; IPR00087; Collagen.
KW Collagen.	RA DR InterPro; IPR02665; P_Rich_Extensn.
SQ SEQUENCE 1802 AA; 172401 MW; 595P16554CBE2D24 CRC64;	RA DR PRINTS; PR01391; Collagen; 1.
Query Match 67.4%; Score 62; DB 5; Length 1802;	RA DR SEQ 224 AA; 24308 MW; F84B9C912D8EC1CD CRC64;
Best Local Similarity 71.4%; Pred. No. 0.65; 3; Mismatches	RA Query Match 66.7%; Score 61; DB 5; Length 224;
Matches 10; Conservative 3; Indels 1; Gaps 0; Gaps 0;	RA Best Local Similarity 66.7%; Pred. No. 0.11; 2; Mismatches 3; Indels 0; Gaps 0; Gaps 0;
Qy 1 GVKGDKGNPGWGA 14	RA Db 118 GGRGDKSHPGMGP 15
Db 1196 GRRGDKSHPGMGP 1209	RA
RESULT 13	RESULT 14
09VWY2 PRELIMINARY; PRT; 224 AA.	Q16593 PRELIMINARY; PRT; 447 AA.
ID 09VWY2	ID 016593
AC	AC 016593:
DT 01-MAY-2000 (TREMBREL 13, Created)	DT 01-NOV-1996 (TREMBREL 01, Created)
DT 01-MAY-2000 (TREMBREL 13, Last sequence update)	DT 01-NOV-1996 (TREMBREL 01, Last sequence update)
DT 01-JUN-2001 (TREMBREL 17, Last annotation update)	DT 01-JUN-2001 (TREMBREL 17, Last annotation update)
DE CG14089 PROTEIN.	DE COLLAGEN-LIKE PROTEIN (47 AA) (FRAGMENT).
GN CG14089.	OS Homo sapiens (Human).
OS Drosophila melanogaster (FRUIT FLY).	OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC Eukaryota; Metazoa; Anthropoda; Tracheata; Hexapoda; Insecta;	OC
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;	OX NEBI_TAXID=227;
OC Ephydriidae; Drosophilidae; Drosophila.	RN
OX NEBI_TAXID=227;	RP SEQUENCE FROM N.A.
RN	RC TISSUE=PLACENTA;
RP SEQUENCE FROM N.A.	RA Kimura S.;
RC STRAIN=BERKELEY;	RL Submitted (FEB-1990) to the EMBL/GenBank/DDBJ databases.
RX MEDLINE=20196006; PubMed=10731132;	RA
RA Adams M.D., Celinski S.E., Holt R.A., Evans C.A., Gocayne J.D.,	RN [2]
RA Ananatis P.G., Scherzer S.E., Li P.W., Hoskins R.A., Galle R.F.,	RP SEQUENCE FROM N.A.
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,	RC TISSUE=PLACENTA;
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,	RA Kimura S.;
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Chapple M., Preiffer B.D.,	RL Submitted (MAY-1992) to the EMBL/GenBank/DDBJ databases.
RA Wan K.H., Doyle C., Baxter G., Heit G., Nelson C.R., Miklos G.L.G.,	DR EMBL; X14963; CAM33085.1; -.
RA Abril J.F., Agbayani A., An H.-J., Andreew-Pfaunhoch C., Baldwin D.,	DR EMBL; X15038; CAM33142.1; -.
RA Balieu R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,	DR InterPro; IPR00087; Collagen.
RA Beeson K.Y., Benos P.V., Bernall B.P., Bhandari D., Bolshakov S.,	DR Pfam; PF01391; Collagen; 5.
RA Borckova D., Botchan M.R., Bouck J., Brokstein P., Brottler P.,	FT NON_TER 1
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra T.,	FT NON_TER 447 447
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,	SQ SEQUENCE 447 AA; 41829 MW; FDB207023DB7CC94 CRC64;
RA de Pablo B., Deicher A., Deng Z., Mayes A.D., Dew I., Dietz S.M.,	Query Match 66.3%; Score 61; DB 4; Length 447;
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,	Best Local Similarity 73.3%; Pred. No. 0.22; 0; Mismatches 4; Indels 0; Gaps 0;
RA Durbin R.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,	Db 280 GOKGDAGNPGDPGTP 294
RA Fosler C., Gabriele A.E., Garg N.S., Geibart W.M., Glasser K.,	RESULT 15
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,	
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,	
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,	
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,	
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,	
RA Lasko P., Lei Y., Levitsky A., Li J., Li Z., Liang Y., Lin X.,	
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,	
RA Q9H0V3	

ID	Q9H0V3	PRELIMINARY;	PRT;	957 AA.	Db	551 GAKGEKGAGCFFGLP 565
AC	Q9H0V3; 01-MAR-2001 (TREMBREL. 16, Created)					: : : : : : :
DT	01-MAR-2001 (TREMBREL. 16, Last sequence update)					
DT	01-DEC-2001 (TREMBREL. 19, Last annotation update)					
DE	HYPOTHETICAL 99.4 KDA PROTEIN.					
GN	DKF2P5641052.					
OS	Homo sapiens (Human).					
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
OX	NCBL-TAXID=9606;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RC	TISSUE=BRAIN;					
RX	MEDLINE=21154917; PubMed=11230166;					
RA	Wiemann S., Wellenreuther R., Gassnerhuber J., Glassl S.,					
RA	Ansorge W., Beecher M., Bloecker H., Bauersachs S., Blum H.,					
RA	Laufer J., Duesterhoeft A., Beyer A., Koehler K., Strack N.,					
RA	Mewes H.W., Oettwelder B., Obermaier B., Tampe J., Heubner D.,					
RA	Wambutt R., Korn B., Klein M., Pousta A.;					
RA	"Towards a Catalog of Human Genes and Proteins: Sequencing and Analysis of 500 Novel Complete Protein Coding Human cDNAs.";					
RL	Genome Res. 11:42-43(2001).					
DR	EMBL; ALI36624; CAB66559.1; -.					
DR	HSSP; P17301; IAOX;					
DR	InterPro; IPR00087; Collagen.					
DR	InterPro; IPR003129; TSN.					
DR	InterPro; IPR002035; vWFa.					
DR	Pfam; PF01391; Collagen; 6.					
DR	Pfam; PF02210; TSPN; 1.					
DR	Pfam; PF00092; vWA; 1.					
DR	PRINTS; PR00053; vWFADOMAIN.					
DR	SMART; SM00210; TSPN; 1.					
DR	SMART; SM0327; vWA; 1.					
DR	PROSITE; PS50234; vWFa; 1.					
KW	Hypothetical protein.					
SQ	SEQUENCE 957 AA; 99413 MW; 391022F715024571 CRC64;					
QY	1 GVKGDKGNGFWPGAP 15					
DB	551 GAKEGKGAGCFFGLP 565					
Query Match	66.3%	Score 61;	DB 4;	Length 957;		
Best Local Similarity	66.7%	Pred. No. 0.48;				
Matches	10;	Conservative 2;	Mismatches 3;	Indels 0;	Gaps 0;	
QY	1 GVKGDKGNGFWPGAP 14					
DB	281 GVKGERGSPGGPA 294					
Query Match	65.2%	Score 60;	DB 13;	Length 310;		
Best Local Similarity	78.6%	Pred. No. 0.21;				
Matches	10;	Conservative 3;	Mismatches 1;	Indels 0;	Gaps 0;	
QY	1 GVKGDKGNGFWPGAP 14					
DB	281 GVKGERGSPGGPA 294					
RESULT 16						
ID	Q96P44	PRELIMINARY;	PRT;	957 AA.		
AC	Q96P44; 01-DEC-2001 (TREMBREL. 19, Created)					
DT	01-DEC-2001 (TREMBREL. 19, Last sequence update)					
DT	01-DEC-2001 (TREMBREL. 19, Last annotation update)					
DE	COLLAGEN XXI.					
GN	COLLAGEN.					
OS	Homo sapiens (Human).					
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
OX	NCBL-TAXID=9606;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RA	Brown A.; Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.					
RL	Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.					
DR	EMBL; ALI38761; CAB66569.1; -.					
DR	InterPro; IPR00087; Collagen.					
DR	Pfam; PF01391; Collagen; 7.					
DR	Pfam; PF01391; Collagen.					
SQ	SEQUENCE 1497 AA; 150419 MW; E01027005F3AE843 CRC64;					
QY	1 GVKGDKGNGFWPGAP 14					
DB	960 GPKGDKGDPGVPGA 973					
Query Match	65.2%	Score 60;	DB 4;	Length 1497;		
Best Local Similarity	78.6%	Pred. No. 1.1;				
Matches	11;	Conservative 1;	Mismatches 2;	Indels 0;	Gaps 0;	
QY	1 GVKGDKGNGFWPGAP 15					

RESULT 19		Db	960 GPKGDKGDFGVPGA 973
Q0MD9	PRELIMINARY;	PRT;	1497 AA.
ID Q0MD9;			
AC Q0MD9;			
DT 01-MAY-2000 (Tremblel. 13, Created)			
DT 01-MAY-2000 (Tremblel. 17, Last annotation update)			
DT 01-JUN-2001 (Tremblel. 17, Last annotation update)			
DE 180 KDA BULLOUS PEMPHIGOID ANTIGEN 2/TYPE XVII COLLAGEN.			
GN BPAG2/COL17A1			
OS Homo sapiens (Human)			
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OC NCBI_TaxID=9606;			
RN [1]			
RP SEQUENCE FROM N.A.; PubMed=9012408;			
RX MEDLINE=97164601; Pubmed=9012408;			
RA Galalica B., Pulkkinen L., Li K., Kuokkanen K., Rynnanen M., McGrath J.A., Uitto J.:			
RT "Cloning of the human type XVII collagen gene (COL17A1), and detection of novel mutations in generalized atrophic benign epidermolysis bullosa"; Am. J. Hum. Genet. 60:352-365(1997).			
RL Am. J. Hum. Genet. 60:352-365(1997).			
DR EMBL; 076604; AAB51499_1; JOINED;			
EMBL; 076565; AAB51499_1; JOINED;			
EMBL; 076566; AAB51499_1; JOINED;			
EMBL; 076567; AAB51499_1; JOINED;			
EMBL; 076568; AAB51499_1; JOINED;			
EMBL; 076569; AAB51499_1; JOINED;			
EMBL; 076570; AAB51499_1; JOINED;			
EMBL; 076571; AAB51499_1; JOINED;			
EMBL; 076572; AAB51499_1; JOINED;			
EMBL; 076573; AAB51499_1; JOINED;			
EMBL; 076574; AAB51499_1; JOINED;			
EMBL; 076575; AAB51499_1; JOINED;			
EMBL; 076576; AAB51499_1; JOINED;			
EMBL; 076577; AAB51499_1; JOINED;			
EMBL; 076578; AAB51499_1; JOINED;			
EMBL; 076579; AAB51499_1; JOINED;			
EMBL; 076580; AAB51499_1; JOINED;			
EMBL; 076581; AAB51499_1; JOINED;			
EMBL; 076582; AAB51499_1; JOINED;			
EMBL; 076583; AAB51499_1; JOINED;			
EMBL; 076584; AAB51499_1; JOINED;			
EMBL; 076585; AAB51499_1; JOINED;			
EMBL; 076586; AAB51499_1; JOINED;			
EMBL; 076587; AAB51499_1; JOINED;			
EMBL; 076588; AAB51499_1; JOINED;			
EMBL; 076589; AAB51499_1; JOINED;			
DR 076590; AAB51499_1; JOINED;			
DR 076591; AAB51499_1; JOINED;			
DR 076592; AAB51499_1; JOINED;			
DR 076593; AAB51499_1; JOINED;			
DR 076594; AAB51499_1; JOINED;			
DR 076595; AAB51499_1; JOINED;			
DR 076596; AAB51499_1; JOINED;			
DR 076597; AAB51499_1; JOINED;			
DR 076598; AAB51499_1; JOINED;			
DR 076599; AAB51499_1; JOINED;			
DR 076600; AAB51499_1; JOINED;			
DR 076601; AAB51499_1; JOINED;			
DR 076602; AAB51499_1; JOINED;			
DR 076603; AAB51499_1; JOINED;			
DR InterPro; IPR000087; Collagen.			
DR Pfam; PF01391; Collagen.			
KW Collagen.			
SQ SEQUENCE 1497 AA; 150459 MW; 9E52B107C6DA6B5C CRC64;			
Query Match 65.2%; Score 60; DB 4; Length 1497;			
Best Local Similarity 78.6%; Pred. No. 1.1; Mismatches 11; Conservative 1; Indels 0; Gaps 0;			
QY 1 GVKGDKGNPGWPGA 14			
RESULT 20		Db	960 GPKGDKGDFGVPGA 973
Q02802	PRELIMINARY;	PRT;	1532 AA.
ID Q02802;			
AC Q02802;			
DT 01-NOV-1996 (Tremblel. 01, Created)			
DT 01-NOV-1996 (Tremblel. 01, Last sequence update)			
DT 01-NOV-1996 (Tremblel. 01, Last sequence update)			
DE AUTOANTIGEN (FRAGMENT).			
GN BP180.			
OS Homo sapiens (Human).			
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OC NCBI_TaxID=9606;			
RN [1]			
RP SEQUENCE FROM N.A.			
RC TISSUE-FORESKIN;			
RX MEDLINE=92381323; Pubmed=1324962;			
RA Gludice G.J., Emery D.J., Diaz L.A.;			
RT "Cloning and Primary structural analysis of the Bullous pemphigoid autotantigen, BP180.", J. Invest. Dermatol. 99:243-250(1992).			
RL DR EMBL; M91660; AAA35605_1; InterPro; IPR000087; Collagen.			
PF Pfam; PF01391; Collagen; 6.			
FT NON-TER 1			
SQ SEQUENCE 1532 AA; 154568 MW; BF35054CF93B8EB3 CRC64;			
Query Match 65.2%; Score 60; DB 4; Length 1532;			
Best Local Similarity 78.6%; Pred. No. 1.1; Mismatches 1; Indels 0; Gaps 0;			
QY 1 GVKGDKGNPGWPGA 14			
Db 995 GPKGDKGDFGVFGA 1008			
RESULT 21		Db	995 GPKGDKGDFGVFGA 1008
O18407	PRELIMINARY;	PRT;	1761 AA.
ID O18407;			
AC O18407;			
DT 01-JAN-1998 (Tremblel. 05, Created)			
DT 01-JAN-1998 (Tremblel. 05, Last sequence update)			
DT 01-JUN-2001 (Tremblel. 17, Last annotation update)			
DE COLLAGEN TYPE IV ALPHA 2.			
GN VKG OR DMCOLA2 OR CG16858.			
OS Drosophila melanogaster (Fruit fly).			
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydriidea; Drosophilidae; Drosophila.			
OC NCBI_TaxID=7277;			
RN [1]			
RP SEQUENCE FROM N.A.			
RA Yasothornsrkul S., Davis W.J., Cramer G., Kimbrell D.A., Dearolf C.R., Submitted (JUL-1996) to the EMBO/GenBank/DBJ databases.			
RL DR EMBL; 065431; AAB54082_1; InterPro; IPR0016075; vkg.			
DR FLYBase; FBgn0016075; vkg.			
DR InterPro; IPR01442; C4.			
DR InterPro; IPR000087; Collagen.			
DR Pfam; PF01391; Collagen; 20.			
DR ProDom; PD003923; C4; 2.			
DR SMART; SM00111; C4; 2.			
SQ SEQUENCE 1761 AA; 175955 MW; FCBA3AFF19121DC6 CRC64;			
Query Match 65.2%; Score 60; DB 5; Length 1761;			
Best Local Similarity 66.7%; Pred. No. 1.3; Mismatches 10; Conservative 2; Indels 3; Gaps 0; Gaps 0;			
QY 1 GVKGDKGNPGWPGA 14			

Db	994	'   : 111111   GRKGETGNPGPGRP 1008	Qy	1	GVKGDKGNPGMPGAP 15     :     :			
<b>RESULT 22</b>								
Q9VMV5		PRELIMINARY;	PRT;	1940 AA.				
ID	Q9VMV5;							
AC	01-MAY-2000 (T-EMBLrel. 13, Created)	ID	063870	PRELIMINARY;	PRT; 2944 AA.			
DT	01-MAY-2000 (T-EMBLrel. 13, Last sequence update)	AC	063870;					
DT	01-JUN-2001 (T-EMBLrel. 17, Last annotation update)	DT	01-NOV-1998 (T-EMBLrel. 08, Last sequence update)					
DE	VKG PROTEIN.	DT	01-DEC-2001 (T-EMBLrel. 19, Last annotation update)					
GN	VKG OR CG16858.	DE	TYPE VII COLLAGEN.					
OS	Drosophila melanogaster (Fruit fly).	GN	COL7A1.					
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;	OS	Mus musculus (Mouse).					
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;	OC	Chordata; Craniata; Vertebrata; Euteleostomi;					
OX	Ephydriidae; Dirosophilidae; Drosophila.	OC	Bukaryota; Metazoa; Rodentia; Sciurognathi; Muridae; Murine; Mus.					
NCBI_TAXID=7227;		OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murine; Mus.					
RN	[1]	OX	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murine; Mus.					
RP	SEQUENCE FROM N.A.	NCBI_TAXID=10080;						
RC	STRAIN=BERKELEY;	RN	[1]					
RX	MEDLINE=20196006; PubMed=10731132;	RP	SEQUENCE FROM N.A.					
RA	Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D., Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F., George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N., Sutton G.J., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X., Brandt R.C., Rogers Y.H.C., Blazier R.G., Champe M., Pfeiffer B.D., Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Miklos G.L.G., Abrial J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D., Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M., Beeson K.Y., Benos P.V., Bernick B.P., Blandari D., Bolshakov S., Borkovka D., Borcman M.R., Bouck J., Brokstein P., Brottier P., Burtis K.C., Busam D.A., Butler H., Cadieu E., Chandra I., Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P., de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M., Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P., Durbin R.K., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W., Fosler C., Gabrieloff A.E., Garg N.S., Gelbart W.M., Glasser K., Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M., Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J., Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C., Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A., Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lal Z., Lasko P., Lei Y., Levitsky D., Li J., Liang Y., Lin X., Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D., Merkulov G., Milashina N.V., Mobarry C., Morris J., Moskrafi A., Mount S.M., Moy M., Murphy B., Murphy L., Muzyk D.M., Nelson D.L., Nelson D.R., Nelson K.A., Nusskern D.R., Paclob J.M., Palazzo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G., Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H., Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T., Spierer E., Spradling A.C., Stapleton M., Strong R., Sun E., Svierskis R., Tector C., Turner R., Venter E., Wang A.H., Wang X., Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissenbach J., Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao O., Zheng L., Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O., Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.; "The genome sequence of Drosophila melanogaster.", Science 287:2185-2195(2000). RML	EMBL; AE003608; AAC02203.1; -.	Qy	1	GVKGDKGNPGMPGAP 15     :     :			
DR	FLYBase; FBgn0016075; vkg.	DR	Q9VMV5; Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Jenkins N.A., Utto J.;	Qy	1	GVKGDKGNPGMPGAP 15     :     :		
DR	InterPro; IPR001442; C4.	DR	"CDNA cloning and chromosomal mapping of the mouse type VII collagen gene (Col7a1): evidence for rapid evolutionary divergence of the gene.",	DR	Q9VMV5; Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Jenkins N.A., Utto J.;	Qy	1	GVKGDKGNPGMPGAP 15     :     :
DR	InterPro; IPR000087; Collagen.	DR	Genomics 16:733-739(1993).	DR	Q9VMV5; Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Jenkins N.A., Utto J.;	Qy	1	GVKGDKGNPGMPGAP 15     :     :
DR	Pfam: PF01413; C4; 2.	DR	[2]	DR	Q9VMV5; Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Jenkins N.A., Utto J.;	Qy	1	GVKGDKGNPGMPGAP 15     :     :
DR	Pfam: PF01391; Collagen; 21.	DR	SEQUENCE FROM N.A.	DR	Q9VMV5; Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Jenkins N.A., Utto J.;	Qy	1	GVKGDKGNPGMPGAP 15     :     :
DR	PROSITE; PS00240; BPTI_KUNITZ_1; 1.	DR	Q9VMV5; Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Jenkins N.A., Utto J.;	DR	Q9VMV5; Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Jenkins N.A., Utto J.;	Qy	1	GVKGDKGNPGMPGAP 15     :     :
DR	PROSITE; PS50279; BPTI_KUNITZ_2; 1.	DR	Q9VMV5; Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Jenkins N.A., Utto J.;	DR	Q9VMV5; Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Jenkins N.A., Utto J.;	Qy	1	GVKGDKGNPGMPGAP 15     :     :
DR	PROSITE; PS50244; WWAFA; 2.	DR	Q9VMV5; Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Jenkins N.A., Utto J.;	DR	Q9VMV5; Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Jenkins N.A., Utto J.;	Qy	1	GVKGDKGNPGMPGAP 15     :     :
DR	PROSITE; PS00327; WVA; 2.	DR	Q9VMV5; Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Jenkins N.A., Utto J.;	DR	Q9VMV5; Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Jenkins N.A., Utto J.;	Qy	1	GVKGDKGNPGMPGAP 15     :     :
DR	PROSITE; PS00092; wva; 2.	DR	Q9VMV5; Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Jenkins N.A., Utto J.;	DR	Q9VMV5; Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Jenkins N.A., Utto J.;	Qy	1	GVKGDKGNPGMPGAP 15     :     :
DR	PRINTS; PRO0759; BASICPTASE.	DR	Q9VMV5; Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Jenkins N.A., Utto J.;	DR	Q9VMV5; Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Jenkins N.A., Utto J.;	Qy	1	GVKGDKGNPGMPGAP 15     :     :
DR	INTERPRO; IPR02223; Kunitz_BPTI.	DR	Q9VMV5; Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Jenkins N.A., Utto J.;	DR	Q9VMV5; Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Jenkins N.A., Utto J.;	Qy	1	GVKGDKGNPGMPGAP 15     :     :
DR	INTERPRO; IPR00235; WWAFA.	DR	Q9VMV5; Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Jenkins N.A., Utto J.;	DR	Q9VMV5; Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Jenkins N.A., Utto J.;	Qy	1	GVKGDKGNPGMPGAP 15     :     :
DR	Pfam: PF01391; Collagen; 22.	DR	Q9VMV5; Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Jenkins N.A., Utto J.;	DR	Q9VMV5; Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Jenkins N.A., Utto J.;	Qy	1	GVKGDKGNPGMPGAP 15     :     :
DR	Pfam: PF00041; fn3; 9.	DR	Q9VMV5; Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Jenkins N.A., Utto J.;	DR	Q9VMV5; Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Jenkins N.A., Utto J.;	Qy	1	GVKGDKGNPGMPGAP 15     :     :
DR	Pfam: PF00014; Kun1; 1.	DR	Q9VMV5; Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Jenkins N.A., Utto J.;	DR	Q9VMV5; Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Jenkins N.A., Utto J.;	Qy	1	GVKGDKGNPGMPGAP 15     :     :
DR	Pfam: PF00092; wva; 2.	DR	Q9VMV5; Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Jenkins N.A., Utto J.;	DR	Q9VMV5; Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Jenkins N.A., Utto J.;	Qy	1	GVKGDKGNPGMPGAP 15     :     :
DR	PRINTS; PRO0453; WWAFA.	DR	Q9VMV5; Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Jenkins N.A., Utto J.;	DR	Q9VMV5; Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Jenkins N.A., Utto J.;	Qy	1	GVKGDKGNPGMPGAP 15     :     :
DR	SMART; SM00060; FN3; 9.	DR	Q9VMV5; Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Jenkins N.A., Utto J.;	DR	Q9VMV5; Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Jenkins N.A., Utto J.;	Qy	1	GVKGDKGNPGMPGAP 15     :     :
DR	SMART; SM00327; WVA; 2.	DR	Q9VMV5; Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Jenkins N.A., Utto J.;	DR	Q9VMV5; Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Jenkins N.A., Utto J.;	Qy	1	GVKGDKGNPGMPGAP 15     :     :
DR	PROSITE; PS00240; BPTI_KUNITZ_1; 1.	DR	Q9VMV5; Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Jenkins N.A., Utto J.;	DR	Q9VMV5; Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Jenkins N.A., Utto J.;	Qy	1	GVKGDKGNPGMPGAP 15     :     :
DR	PROSITE; PS50279; BPTI_KUNITZ_2; 1.	DR	Q9VMV5; Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Jenkins N.A., Utto J.;	DR	Q9VMV5; Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Jenkins N.A., Utto J.;	Qy	1	GVKGDKGNPGMPGAP 15     :     :
DR	PROSITE; PS50244; WWAFA; 2.	DR	Q9VMV5; Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Jenkins N.A., Utto J.;	DR	Q9VMV5; Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Jenkins N.A., Utto J.;	Qy	1	GVKGDKGNPGMPGAP 15     :     :
DR	PROSITE; PS00240; BPTI_KUNITZ_1; 1.	DR	Q9VMV5; Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Jenkins N.A., Utto J.;	DR	Q9VMV5; Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Jenkins N.A., Utto J.;	Qy	1	GVKGDKGNPGMPGAP 15     :     :
DR	PROSITE; PS50279; BPTI_KUNITZ_2; 1.	DR	Q9VMV5; Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Jenkins N.A., Utto J.;	DR	Q9VMV5; Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Jenkins N.A., Utto J.;	Qy	1	GVKGDKGNPGMPGAP 15     :     :
DR	PROSITE; PS50244; WWAFA; 2.	DR	Q9VMV5; Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Jenkins N.A., Utto J.;	DR	Q9VMV5; Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Jenkins N.A., Utto J.;	Qy	1	GVKGDKGNPGMPGAP 15     :     :
DR	PROSITE; PS50244; WWAFA; 2.	DR	Q9VMV5; Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Jenkins N.A., Utto J.;	DR	Q9VMV5; Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Jenkins N.A., Utto J.;	Qy	1	GVKGDKGNPGMPGAP 15     :     :
DR	PROSITE; PS50244; WWAFA; 2.	DR	Q9VMV5; Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Jenkins N.A., Utto J.;	DR	Q9VMV5; Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Jenkins N.A., Utto J.;	Qy	1	GVKGDKGNPGMPGAP 15     :     :
DR	PROSITE; PS50244; WWAFA; 2.	DR	Q9VMV5; Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Jenkins N.A., Utto J.;	DR	Q9VMV5; Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Jenkins N.A., Utto J.;	Qy	1	GVKGDKGNPGMPGAP 15     :     :
DR	PROSITE; PS50244; WWAFA; 2.	DR	Q9VMV5; Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Jenkins N.A., Utto J.;	DR	Q9VMV5; Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Jenkins N.A., Utto J.;	Qy	1	GVKGDKGNPGMPGAP 15     :     :
DR	PROSITE; PS50244; WWAFA; 2.	DR	Q9VMV5; Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Jenkins N.A., Utto J.;	DR	Q9VMV5; Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Jenkins N.A., Utto J.;	Qy	1	GVKGDKGNPGMPGAP 15     :     :
DR	PROSITE; PS50244; WWAFA; 2.	DR	Q9VMV5; Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Jenkins N.A., Utto J.;	DR	Q9VMV5; Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Jenkins N.A., Utto J.;	Qy	1	GVKGDKGNPGMPGAP 15     :     :
DR	PROSITE; PS50244; WWAFA; 2.	DR	Q9VMV5; Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Jenkins N.A., Utto J.;	DR	Q9VMV5; Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Jenkins N.A., Utto J.;	Qy	1	GVKGDKGNPGMPGAP 15     :     :
DR	PROSITE; PS50244; WWAFA; 2.	DR	Q9VMV5; Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Jenkins N.A., Utto J.;	DR	Q9VMV5; Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Jenkins N.A., Utto J.;	Qy	1	GVKGDKGNPGMPGAP 15     :     :
DR	PROSITE; PS50244; WWAFA; 2.	DR	Q9VMV5; Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Jenkins N.A., Utto J.;	DR	Q9VMV5; Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Jenkins N.A., Utto J.;	Qy	1	GVKGDKGNPGMPGAP 15     :     :
DR	PROSITE; PS50244; WWAFA; 2.	DR	Q9VMV5; Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Jenkins N.A., Utto J.;	DR	Q9VMV5; Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Jenkins N.A., Utto J.;	Qy	1	GVKGDKGNPGMPGAP 15     :     :
DR	PROSITE; PS50244; WWAFA; 2.	DR	Q9VMV5; Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Jenkins N.A., Utto J.;	DR	Q9VMV5; Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Jenkins N.A., Utto J.;	Qy	1	GVKGDKGNPGMPGAP 15     :     :
DR	PROSITE; PS50244; WWAFA; 2.	DR	Q9VMV5; Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Jenkins N.A., Utto J.;	DR	Q9VMV5; Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Jenkins N.A., Utto J.;	Qy	1	GVKGDKGNPGMPGAP 15     :     :
DR	PROSITE; PS50244; WWAFA; 2.	DR	Q9VMV5; Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Jenkins N.A., Utto J.;	DR	Q9VMV5; Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Jenkins N.A., Utto J.;	Qy	1	GVKGDKGNPGMPGAP 15     :     :
DR	PROSITE; PS50244; WWAFA; 2.	DR	Q9VMV5; Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Jenkins N.A., Utto J.;	DR	Q9VMV5; Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Jenkins N.A., Utto J.;	Qy	1	GVKGDKGNPGMPGAP 15     :     :
DR	PROSITE; PS50244; WWAFA; 2.	DR	Q9VMV5; Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Jenkins N.A., Utto J.;	DR	Q9VMV5; Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Li K., Christiano A.M., Copeland N.G., Gilbert D.J., Chu M.L., Jenkins N.A., Utto J.;	Qy	1	GVKGDKGNPGMPGAP 15     :     :
DR	PROSITE; PS50244; WWAFA; 2.	DR	Q9VMV5; Li K., Christiano A					

AC Q26312;  
 DT 01-NOV-1996 (TREMBREL. 01, Created)  
 DT ID 09JUC7 PRELIMINARY; PRT; 771 AA.  
 DT 01-NOV-1996 (TREMBREL. 01, Last sequence update)  
 DE 09JUC7; AC 09JUC7;  
 DE NONFIBRILLAR COLLAGEN PROTEIN (FRAGMENT).  
 OS STRONGYLOCENTROTUS PURPURATUS (purple sea urchin).  
 OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;  
 OC Echinida; Euechinida; Echinacea; Echinolla; Strongylocentrotidae;  
 OC Strongylocentrotus.  
 OX NCBI\_TAXID=7668;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92038439; PubMed=1936564;  
 RA Wessel G.M., Etkin M., Benson S.;  
 RT "Primary mesenchyme cells of the sea urchin embryo require an  
 autonomously produced, nonfibrillar collagen for spiculogenesis.";  
 RL Dev. Biol. 148:261-272(1991).  
 DR EMBL; S64572; AAC020970.1;  
 DR InterPro; IPR000087; Collagen.  
 PRfam; PF01391; Collagen; 13.  
 KW Collagen.  
 FT NON-TER 1  
 SQ SEQUENCE 907 AA; '85400 MW; B8ECBF6FDAA59D88 CRC64;  
 Query Match 64.7%; Score 59.5; DB 5; Length 907;  
 Best Local Similarity 54.2%; Pred. No. 0.77;  
 Matches 13; Conservative 0; Mismatches 2; Indels 9; Gaps 1;  
 QY 1 GVKGDD-----KGNRGWPGAP 15  
 Db 546 GVKGDPGRTGPEGAKGNPGLPGIP 569  
 RESULT 25  
 007265 PRELIMINARY; PRT; 1752 AA.  
 AC 007265; ID 007265 PRELIMINARY; PRT; 1752 AA.  
 AC 007265; ID 007265 PRELIMINARY; PRT; 1752 AA.  
 DT 01-NOV-1996 (TREMBREL. 01, Created)  
 DT ID 09JUC7 PRELIMINARY; PRT; 771 AA.  
 DT 01-NOV-1996 (TREMBREL. 01, Last sequence update)  
 DE 09JUC7; AC 09JUC7;  
 DE COLPALARNA.  
 OS STRONGYLOCENTROTUS PURPURATUS (Purple sea urchin).  
 OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;  
 OC Echinida; Euechinida; Echinacea; Echinolla; Strongylocentrotidae;  
 OC Strongylocentrotus.  
 OX NCBI\_TAXID=7668;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93186842; PubMed=8444899;  
 RA Exposito J.-Y., Di Liberto M., Ramirez F.;  
 RT "Complete primary structure of a sea-urchin type IV collagen and  
 analysis of the 5' end of its gene.",  
 J. Biol. Chem. 268:5243-5244(1993).  
 DR EMBL; L02917; AAC30039.1; -.  
 DR InterPro; IPR001442; C4.  
 DR InterPro; IPR000087; Collagen.  
 PRfam; PF01413; C4; 2.  
 PRfam; PF01391; Collagen; 22.  
 PRfam; PD003923; C4; 2.  
 SMART; SM00111; C4; 2.  
 KW Collagen.  
 SQ SEQUENCE 1752 AA; 170210 MW; 1AE5AAA21569346D CRC64;  
 Query Match 64.7%; Score 59.5; DB 5; Length 1752;  
 Best Local Similarity 54.2%; Pred. No. 1.5;  
 Matches 13; Conservative 0; Mismatches 2; Indels 9; Gaps 1;  
 QY 1 GVKGDD-----KGNRGWPGAP 15  
 Db 1177 GVKGDPGRTGPEGAKGNPGLPGIP 1200  
 RESULT 26  
 09JUC7 PRELIMINARY; PRT; 771 AA.  
 ID 09JUC7 PRELIMINARY; PRT; 771 AA.  
 AC 09JUC7; AC 09JUC7;  
 DE 01-MAY-2000 (TREMBREL. 13, Created)  
 DT 01-MAY-2000 (TREMBREL. 13, Last sequence update)  
 DE COIA5 (COLLAGEN ALPHA 5(IV) CHAIN PRECURSOR) (FRAGMENT).  
 GN COIA5.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Primates; Catarrhini; Hominoidea; Homo.  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 OX NCBI\_TAXID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 ID 09JUC7 PRELIMINARY; PRT; 771 AA.  
 AC 09JUC7; AC 09JUC7;  
 RL Submitted (JAN-1999) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AL031622; CAA20937.1; -.  
 DR InterPro; IPR000087; Collagen.  
 PRfam; PF01391; Collagen; 10.  
 KW Collagen.  
 FT NON-TER 1  
 FT NON-TER 771 AA; 771 MW; 9CECB0373E545A61 CRC64;  
 SQ SEQUENCE 771 AA; 72760 MW; 9CECB0373E545A61 CRC64;  
 Query Match 64.1%; Score 59; DB 4; Length 771;  
 Best Local Similarity 73.3%; Pred. No. 0.78;  
 Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 GVKGDKGNPGWPGAP 15  
 Db 528 GMKGDKGELGSPGAP 542  
 RESULT 27  
 09JUC4 PRELIMINARY; PRT; 1431 AA.  
 ID 09JUC4 PRELIMINARY; PRT; 1431 AA.  
 AC 09JUC4; ID 09JUC4 PRELIMINARY; PRT; 1431 AA.  
 AC 09JUC4; ID 09JUC4 PRELIMINARY; PRT; 1431 AA.  
 DT 01-OCT-2000 (TREMBREL. 15, Created)  
 DT 01-OCT-2000 (TREMBREL. 15, Last sequence update)  
 DE 01-JUN-2001 (TREMBREL. 17, Last annotation update)  
 DE COLLAGEN TYPE XVII.  
 OS Mesocricetus auratus (Golden hamster).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;  
 OC Mesocricetus  
 OX NCBI\_TAXID=10036;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Yamamoto K., Inoue N., Fujimori A., Saito T., Shinkai H., Sakiyama H.;  
 RT "Mosoicrictetus auratus mRNA for type XVII collagen.",  
 RL Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AB027159; BAB49430.1; -.  
 DR InterPro; IPR000087; Collagen.  
 PRfam; PF01391; Collagen; 5.  
 SQ SEQUENCE 1431 AA; 144579 MW; 4315631FEB2C9A5C CRC64;  
 Query Match 64.1%; Score 59; DB 11; Length 1431;  
 Best Local Similarity 60.7%; Pred. No. 1.5;  
 Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 GVKGDKGNPGWPGAP 15  
 Db 945 GPKGDPGQGPQGPCTP 959  
 RESULT 28  
 007563 PRELIMINARY; PRT; 1433 AA.  
 ID 007563 PRELIMINARY; PRT; 1433 AA.  
 AC 007563; AC 007563;  
 DT 01-NOV-1996 (TREMBREL. 01, Created)  
 DT 01-NOV-1996 (TREMBREL. 01, Last sequence update)

DE PROCOLLAGEN, TYPE XVII, ALPHA 1 (BULLOUS PEMPHIGOID AUTOANTIGEN  
DE BP180) (COLLAGEN XVII).  
DE COL17AI.  
GN MUS musculus (Mouse).  
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TAXID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BALB/C;  
RX MEDLINE=93232041; PubMed=8473327;  
RA Li K., Tamai K., Iain E.M., Ultto J.;  
RT "Cloning of type XVII collagen. Complementary and genomic DNA  
sequences of mouse 180-kilodalton bullous pemphigoid antigen (BPAG2)  
predict an interrupted collagenous domain, a transmembrane segment,  
and unusual features in the 5'-end of the gene and the 3'-untranslated  
region of the mRNA.";  
RT J. Biol. Chem. 268:8825-8834(1993).  
RT -!!- FUNCTION: THE COLLAGEN DOMAINS OF BP180 ALLOW THE HEMIDESMOSOME TO  
FORM STABLE INTERACTIONS WITH THE CONSTITUENTS OF THE EXTRACELLULAR  
MATRIX OF THE CUTANEOUS BASEMENT MEMBRANE ZONE.  
CC -!!- SUBCELLULAR LOCATION: A COMPONENT OF THE HEMIDESMOSOME.  
CC -!!- DISEASE: UPON DISRUPTION OF BP180 BY AUTOANTIBODIES THE  
HEMIDESMOSOME IS DISRUPTED LEADING TO THE BLISTERING SKIN DISORDER  
CC BULLOUS PEMPHIGOID (BP).  
DR EMBL; LO0407; AAA37433; 1; -.  
DR MGD; MGI:88450; Col17ai.  
DR InterPro; IPR000087; Collagen.  
DR Pfam; PF01391; Collagen; 5.  
KW Antigen; Cell adhesion.  
SEQUENCE 1433 AA; 144087 MW; BBF808832A19922C CRC64;  
Query Match 64.1%; Score 59; DB 11; Length 1433;  
Best Local Similarity 66.7%; Pred. No. 1.5;  
Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
QY 1 GVKGDKGNPGWGP 15  
| | |||:||| |||  
Db 948 GPKDQGDPGVPGP 962

RESULT 30  
ID Q9J104 PRELIMINARY; PRT; 1737 AA.  
AC Q9J104;  
AC Q9J104; Created)  
DT 01-OCT-2000 (TREMBIREL. 15, Last sequence update)  
DT 01-OCT-2000 (TREMBIREL. 15, Last sequence update)  
DE ALPHA 4 TYPE V COLLAGEN.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TAXID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SPRAGUE DAWLEY;  
RX MEDLINE=20428740; PubMed=10052920;  
RA Chernousov M.A., Rothblum K., Tyler W.A., Stahl R.C., Carey D.J.;  
RT "Schwann cells synthesize type V collagen that contains a novel alpha  
4 chain. Molecular cloning, biochemical characterization, and high  
affinity heparin binding of alpha(V) collagen.";  
RT J. Biol. Chem. 275:28208-28215(2000).  
DR EMBL; AF272661; AAC76432; 1; -.  
DR InterPro; IPR000087; Collagen.  
DR InterPro; IPR000885; Fib\_collagen\_C.  
DR InterPro; IPR001791; Laminin\_G.  
DR Prodom; PD002078; Fib\_collagen\_C; 1.  
DR SMART; SM0038; Colf1; 1.  
DR SMART; SM00282; Lamc; 1.  
DR SMART; SM00210; TSPN; 1.  
KW Collagen.  
SQ SEQUENCE 1737 AA; 171574 MW; D635D5D57481C257 CRC64;

RESULT 31  
ID Q9Q2R9 PRELIMINARY; PRT; 1682 AA.  
AC Q9Q2R9;  
DT 01-MAY-2000 (TREMBIREL. 13, Created)  
DT 01-JUN-2001 (TREMBIREL. 17, Last annotation update)  
DE ALPHA 4 COLLAGEN IV.  
GN COLA4.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBL\_TAXID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-KIDNEY;  
RX MEDLINE=2005934; PubMed=10534397;  
RA Lu W., Phillips C.L., Killen P.D., Hlaing T., Harrison W.R.,  
RA Elder F.F.B., Miner J.H., Overbeek P.A., Meisler M.H.;  
RT "Insertional mutation of the collagen genes col4a3 and col4a4 in a  
mouse model of alport syndrome.";  
RL Genomics 61:113-124(1999).  
DR EMBL; AF169388; AAC05450; 1; -.  
DR MGI; MGI:104687; Col4a4.  
DR InterPro; IPR01442; C4;  
DR InterPro; IPR01413; C4; 2.  
DR Pfam; PF01391; Collagen; 21.  
DR Prodom; PD003923; C4; 2.  
DR SMART; SM00111; C4; 2.  
KW Collagen.  
SQ SEQUENCE 1682 AA; 164096 MW; 6F7B679EDD76E904 CRC64;

Query Match 64.1%; Score 59; DB 11; Length 1682;  
Best Local Similarity 60.0%; Pred. No. 1.7; Matches 9; Conservative 4; Mismatches 2; Indels 0; Gaps 0;  
QY 1 GVKGDKGNPGWGP 15  
| :|||:||| |||  
Db 176 GIQQDRGDPGPPGLP 190

Query Match 64.1%; Score 59; DB 11; Length 1737;  
Best Local Similarity 76.9%; Pred. No. 1.8; Matches 10; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
QY 1 GVKGDKGNPGWGP 13  
| |||:||| |||  
Db 740 GPKDQRGNPGLPG 752

Search completed: November 1, 2002, 12:53:44  
Job time : 24.5 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

On protein - protein search, using SW model.  
Run on: November 1, 2002, 12:14:41. ; Search time 26.5 Seconds

(without alignments) updates/sec  
62.872 Million cell

Title: US-09-529-691A-3  
Perfect score: 92  
Sequence: 1 PAGPGPGNGKDGKVG 15  
Scoring table: BL05UM62  
Gapop 10.0 , Gapext 0.5  
Searched: 74574 seqs, 11073796 residues  
Total number of hits satisfying chosen parameters: 74574  
Minimum DB seq length: 0  
Maximum DB seq length: 200000000  
Post-processing: Minimum Match 0%  
Listing first 45 summaries

**Database :**

1: /SIDS1/gcdata/geneseq/geneseqp-emb1/AA1980.DAT:\*

2: /SIDS1/gcdata/geneseq/geneseqp-emb1/AA1981.DAT:\*

3: /SIDS1/gcdata/geneseq/geneseqp-emb1/AA1982.DAT:\*

4: /SIDS1/gcdata/geneseq/geneseqp-emb1/AA1983.DAT:\*

5: /SIDS1/gcdata/geneseq/geneseqp-emb1/AA1984.DAT:\*

6: /SIDS1/gcdata/geneseq/geneseqp-emb1/AA1985.DAT:\*

7: /SIDS1/gcdata/geneseq/geneseqp-emb1/AA1986.DAT:\*

8: /SIDS1/gcdata/geneseq/geneseqp-emb1/AA1987.DAT:\*

9: /SIDS1/gcdata/geneseq/geneseqp-emb1/AA1988.DAT:\*

10: /SIDS1/gcdata/geneseq/geneseqp-emb1/AA1989.DAT:\*

11: /SIDS1/gcdata/geneseq/geneseqp-emb1/AA1990.DAT:\*

12: /SIDS1/gcdata/geneseq/geneseqp-emb1/AA1991.DAT:\*

13: /SIDS1/gcdata/geneseq/geneseqp-emb1/AA1992.DAT:\*

14: /SIDS1/gcdata/geneseq/geneseqp-emb1/AA1993.DAT:\*

15: /SIDS1/gcdata/geneseq/geneseqp-emb1/AA1994.DAT:\*

16: /SIDS1/gcdata/geneseq/geneseqp-emb1/AA1995.DAT:\*

17: /SIDS1/gcdata/geneseq/geneseqp-emb1/AA1996.DAT:\*

18: /SIDS1/gcdata/geneseq/geneseqp-emb1/AA1997.DAT:\*

19: /SIDS1/gcdata/geneseq/geneseqp-emb1/AA1998.DAT:\*

20: /SIDS1/gcdata/geneseq/geneseqp-emb1/AA1999.DAT:\*

21: /SIDS1/gcdata/geneseq/geneseqp-emb1/AA2000.DAT:\*

22: /SIDS1/gcdata/geneseq/geneseqp-emb1/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Match Length	DB ID	Description
1	63	68.5	534 18 AAW12844	RESULT 1 ID AAW12844 standard; Peptide: 534 AA.
2	63	68.5	535 18 AAW12841	XX AC AAW12844;
3	63	68.5	537 18 AAW12840	XX DT 15-DEC-1997 (first entry)
4	63	68.5	674 21 AAB53439	XX DE Pro-alpha2(I); (III)BGR chimeric protein.
5	63	68.5	1040 21 AAY84547	XX KW C-Propeptid; recognition sequence; procollagen; monomer chain; therapy; KW trimersised pro-alpha chain; fibril; procollagen suicide; wound healing; KW fibrotic disease; human; chimeric protein.
6	63	68.5	1040 21 AAY84548	XX OS Home sapiens.
7	63	68.5	1366 16 AAY1702	XX PN WO9708311-A1.
8	63	68.5	1366 21 AAY96123	XX PD 06-MAR-1997.
9	63	68.5	1366 21 AYX56801	XX PF 30-AUG-1996; 95NO-GB02122.
10	63	68.5	1366 22 ABP50293	XX PR 14-JUN-1996; 96GB-0012476.
11	63	68.5	1366 22 AAE02536	PR 31-AUG-1995; 95CB-001773.
				PR 23-MAR-1996; 96GB-0006152.
				PA (UYMA-) UNIV VICTORIA MANCHESTER.
				XX PI Bulleid N, Kandler K;
				XX DR WPT; 1997-179268/16.
				XX PT Novel pro:collagen mol. - comprising attached to an alien collagen alpha-chain or non-collagen material,

PT useful e.g. for wound healing  
 XX  
 PS Example 1; Page 38-41; 69pp; English.

CC This sequence represents a chimeric procollagen molecule of the  
 CC pro-alpha<sub>1</sub>(II) chain inserted into the pro-alpha<sub>2</sub>(I) chain sequence. The  
 CC C-propeptide is implicated in the assembly of the monomer chains into  
 CC trimerised pro-alpha chains prior to cleavage of the N- and C-propeptides  
 CC and formation of collagen in fibril-forming pro-alpha chains. The  
 CC C-propeptides determine the type-specific assembly of the moieties to  
 CC which they are attached. The molecule of the invention comprises a first  
 CC moiety having procollagen C-propeptide activity attached to a second  
 CC moiety, which is an alien collagen alpha-chain or a non-collagen  
 CC material. The novel collagen molecule can be used for treatment or  
 CC diagnosis in humans or animals, especially for the treatment of  
 CC procollagen suicide, as an adhesive or implant, to promote (chronic)  
 CC wound healing or fibrotic diseases with reduced scarring or for use in  
 CC photography, brewing, foodstuffs or textiles. The novel collagen  
 CC molecules, especially when containing substitutions in the recognition  
 CC site, may have significantly altered properties and characteristics, such  
 CC as different binding kinetics or alpha-chain selection properties.  
 XX

SQ Sequence 534 AA;

Query	Match	Score	DB	Length
Best Local Matches	Similarity	68.5%	18	534
11; Conservative	Pred. No.	0.65		
	Mismatches	3		
	Indels	0		
	Gaps	0		

OY 1 PAGPMPNGKDGKVG 15  
 Db 223 PAGPSGPAGKDGRTG 237

RESULT 2

ID	AAW12841	Standard; Protein; 535 AA.
XX	AAW12841;	
AC		
XX	15-DEC-1997 (first entry)	
DT		
XX	Truncated pro-alpha <sub>2</sub> (I) chain.	
DE		
XX	C-propeptide; recognition sequence; procollagen; monomer chain; therapy; trimerised pro-alpha chain; fibril; procollagen suicide; wound healing; fibrotic disease; human.	
XX	Homo sapiens.	
OS		
XX	WO9708311-A1.	
PN		
XX	06-MAR-1997.	
PD		
XX	30-AUG-1996; 96WO-GB02122.	
PF		
XX	14-JUN-1996; 96GB-0012476.	
PR		
PR	31-AUG-1995; 95GB-0017773.	
PR		
XX	23-MAR-1996; 96GB-0006152.	
XX		
PF		
XX	30-AUG-1996; 96WO-GB02122.	
PR		
PR	14-JUN-1996; 96GB-0012476.	
PR		
PR	31-AUG-1995; 95GB-001773.	
PR		
XX	23-MAR-1996; 96GB-0006152.	
XX		
PA	(UYMA-) UNIV VICTORIA MANCHESTER.	
XX		
PI	Bulleid N, Kadler K;	
XX		
DR	WPI: 1997-17928/16.	
DR	N-PSDB; AAT59891.	
XX		
PT	Novel pro-collagen mol. - comprising pro:collagen C-pro:peptide attached to an alien collagen alpha-chain or non-collagen material, useful e.g. for wound healing	
PT		
XX		
PS	Example 1; Page 25-28; 69pp; English.	

CC This sequence represents a truncated procollagen pro-alpha<sub>2</sub>(I) chain that  
 CC can be used in the procollagen molecules of the invention. The  
 CC C-propeptide is implicated in the assembly of the monomer chains into  
 CC trimerised pro-alpha chains prior to cleavage of the N- and C-propeptides  
 CC and formation of collagen in fibril-forming pro-alpha chains. The  
 CC C-propeptides determine the type-specific assembly of the moieties to  
 CC which they are attached. The molecule of the invention comprises a first  
 CC moiety having procollagen C-propeptide activity attached to a second  
 CC moiety, which is an alien collagen alpha-chain or a non-collagen  
 CC material. The novel collagen molecule can be used for treatment or  
 CC diagnosis in humans or animals, especially for the treatment of  
 CC procollagen suicide, as an adhesive or implant, to promote (chronic)  
 CC wound healing or fibrotic diseases with reduced scarring or for use in  
 CC photography, brewing, foodstuffs or textiles. The novel collagen  
 CC molecules, especially when containing substitutions in the recognition  
 CC site, may have significantly altered properties and characteristics, such  
 CC as different binding kinetics or alpha-chain selection properties.  
 XX

SQ Sequence 535 AA;

Query	Match	Score	DB	Length
Best Local Matches	Similarity	68.5%	18	535
11; Conservative	Pred. No.	0.65		
	Mismatches	3		
	Indels	0		
	Gaps	0		

OY 1 PAGPMPNGKDGKVG 15  
 Db 224 PAGPSGPAGKDGRTG 238

RESULT 3

ID	AAW12840	Standard; Protein; 537 AA.
XX	AAW12840;	
AC		
XX	15-DEC-1997 (first entry)	
DT		
XX	Pro-alpha <sub>2</sub> (I): (III) CP chimeric protein.	
DE		
XX	C-propeptide; recognition sequence; procollagen; monomer chain; therapy; trimerised pro-alpha chain; fibril; procollagen suicide; wound healing; fibrotic disease; human; chimeric protein.	
XX	Homo sapiens.	
OS		
XX	WO9708311-A1:	
PN		
XX	06-MAR-1997.	
PD		
XX	30-AUG-1996; 96WO-GB02122.	
PF		
XX	14-JUN-1996; 96GB-0012476.	
PR		
PR	31-AUG-1995; 95GB-001773.	
PR		
XX	23-MAR-1996; 96GB-0006152.	
XX		
PA	(UYMA-) UNIV VICTORIA MANCHESTER.	
XX		
PI	Bulleid N, Kadler K;	
XX		
DR	WPI: 1997-17928/16.	
DR	N-PSDB; AAT59891.	
XX		
PT	Novel pro-collagen mol. - comprising pro:collagen C-pro:peptide attached to an alien collagen alpha-chain or non-collagen material, useful e.g. for wound healing	
PT		
XX		
PS	Claim 16; Page 35-38; 69pp; English.	

CC This sequence represents a chimeric procollagen molecule of the  
 CC invention. This sequence has the procollagen C-prepropeptide from the  
 CC pro-alpha<sub>1</sub>(II) chain attached to the pro-alpha<sub>2</sub>(I) chain sequence. The  
 CC C-propeptide is implicated in the assembly of the monomer chains into  
 CC trimerised proalpha chains prior to cleavage of the N- and C-propeptides  
 CC and formation of collagen in fibril-forming pro-alpha chains. The

CC C-propeptides determine the type-specific assembly of the moieties to which they are attached. The molecule of the invention comprises a first CC moiety having procollagen C-propeptide activity attached to a second CC moiety, which is an alien collagen alpha-chain or a non-collagen material. The novel collagen molecule can be used for treatment or CC diagnosis in humans or animals, especially for the treatment of CC procollagen suicide, as an adhesive or implant, to promote (chronic) wound healing or fibrotic diseases with reduced scarring or for use in CC photography, brewing, foodstuffs or textiles. The novel collagen molecules, especially when containing substitutions in the recognition CC site, may have significantly altered properties and characteristics, such CC as different binding kinetics or alpha-chain selection properties.

XX SQ Sequence 537 AA;

Query Match 68.5%; Score 63; DB 18; Length 537;  
Best Local Similarity 73.3%; Pred. No. 0.65; Matches 11; Mismatches 1; Indels 3; Gaps 0;

Oy 1 PAGPMCPNGKDGKV 15  
Db 224 PAGPSGCPAGKDGRTG 238

RESULT 4

ID AAB53439  
XX AAB53439 standard; Protein: 674 AA.  
AC ;  
XX AAB53439;  
DT 09-MAR-2001 (first entry)  
DE Human colon cancer antigen protein sequence SEQ ID NO:979.  
KW identification; cytostatic; cardioactive; neuroprotective; vulneary; immunomodulatory; muscular; gynaecological; gastrointestinal; detection; nephrotoxic; antiinfective; antibacterial; gene therapy; wound; neural disorder; immune system disorder; muscular disorder; reproductive disorder; gastrointestinal disorder; renal disorder; infectious disease; cardiovascular disorder.  
XX OS Homo sapiens.  
XX PN WO200055351-A1.  
XX PD 21-SEP-2000.  
XX PR 08-MAR-2000; 2000WO-US05983.  
XX PR 12-MAR-1999; 99US-0124270.  
XX PA (HUMA-) HUMAN GENOME SCI INC.  
PT Rosen CA, Ruben SM;  
XX DR WPI; 2000-587534/55.  
DR N-PSDB; AAC98195.  
XX Colon cancer associated gene sequences, referred to as colon cancer PT antigens, useful for the treatment, prevention, and diagnosis of colon disorders such as colon cancer -  
PS Claim 11; Page 1551-1553; 210pp; English.

XX AAC97991 to AAC98763 encode the human colon cancer associated proteins, called human colon cancer antigens, given in AAB53434 to AAB54005. The human colon cancer antigens can have cytostatic, cardioactive, muscular, neuroprotective, immunomodulatory, gynaecological, gastrointestinal, vulnerable, nephrotoxic, antiinfective and antibacterial activities, and can be used in gene therapy. The colon cancer antigen polynucleotides, proteins and antibodies to the proteins are useful for the prevention, treatment and diagnosis of colon disorders, such as colon cancer. The

CC polyribonucleotides may be used in diagnostics and research, such as for CC chromosome identification, and as hybridisation probes. The proteins CC may also be used to prevent diseases such as neural disorders, immune CC system disorders, muscular disorders, reproductive disorders, gastrointestinal disorders, wounds, renal disorders, Infectious CC diseases, and cardiovascular disorders. AAC9764 to AAC9872 and CC AAC98007 represent sequences used in the exemplification of the present invention.

XX SQ Sequence 674 AA;

Query Match 68.5%; Score 63; DB 21; Length 674;  
Best Local Similarity 73.3%; Pred. No. 0.82; Matches 11; Mismatches 1; Indels 3; Gaps 0;

Oy 1 PAGPMCPNGKDGKV 15  
Db 363 PAGPSGCPAGKDGRTG 377

RESULT 5

ID AAY84547  
XX AAY84547 standard; Protein: 1040 AA.  
AC ;  
XX AAY84547;  
DT 25-JUL-2000 (first entry)  
DE A human collagen 1 (alpha2) protein helical region.  
KW Extracellular matrix protein; self aggregation; hydroxylated proline; trans-4-hydroxyproline; 3-hydroxyproline; recombinant protein production; collagen; fibrinogen; fibronectin; post translational hydroxylation.  
XX OS Homo sapiens.  
XX PN EP992586-A2.  
XX PD 12-APR-2000.  
XX PF 07-OCT-1999; 99EP-0119184.  
XX PR 09-OCT-1998; 98US-0169768.  
XX PA (USSU ) US SURGICAL CORP.  
XX PI Gruskin EA, Buechter DD, Zhang G, Connolly K;  
XX DR WPI; 2000-259138/23.  
XX N-PSDB; AAA12510.  
PT Production of extracellular matrix proteins containing 4-trans-hydroxyproline results in native self aggregating proteins, useful on medical implants -  
PT Example 14; Fig 49A-E; 260pp; English.

XX PS The specification describes a method for producing an extracellular matrix protein or its fragment. The extracellular matrix protein is capable of self aggregating in a cell which does not ordinarily hydroxylate prolines. The method comprises optimising a nucleic acid sequence for expression in the cell by substitution of codons preferred by that cell for naturally occurring codons not preferred by the cell; incorporating the nucleic acid sequence into the cell; and contacting the cell with a hypertonic growth medium containing at least one amino acid selected from the group consisting of trans-4-hydroxyproline and 3-hydroxyproline to allow at least one of the amino acids to be assimilated into the cell and incorporated into the extracellular matrix protein. The method may be used to make host cells assimilate and incorporate trans-4-hydroxyproline into proteins. This is especially useful in the recombinant production of proteins such as collagen, fibrinogen and fibronectin whose ability to self aggregate and produce functional proteins depends on the post-translational hydroxylation of

CC proline. The method is also useful in studying the structure and function of polypeptides which do not normally contain trans- $\alpha$ -hydroxyproline. The present sequence represents human collagen 1 ( $\alpha_2$ ) helical region, which may be produced using the method of the invention.

XX	Sequence	1040 AA:	Query Match	68.5%; Score 63; DB 21; Length 1040;
CC	Best Local Similarity	73.3%; Pred. No. 1.3;	Matches	11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
CC	Matches	11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;	Qy	1 PAGPWGPNGDKGKG 15
CC	Db	976 PAGPSGPAGKDGRIG 990	Qy	1 PAGPWGPNGDKGKG 15
			Db	976 PAGPSGPAGKDGRIG 990
RESULT 6	AAY4548	ID AAY4548 standard; Protein: 1040 AA.	Query Match	68.5%; Score 63; DB 21; Length 1040;
XX	AC	AAY4548;	Matches	11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
XX	DT	25-JUL-2000 (first entry)	Qy	1 PAGPWGPNGDKGKG 15
XX	DE	A human collagen 1 ( $\alpha_2$ ) protein helical region.	Db	976 PAGPSGPAGKDGRIG 990
XX	Extracellular matrix protein; self aggregation; hydroxylated proline; recombinant protein production; trans-4-hydroxyproline; 3-hydroxyproline; collagen; fibrinogen; fibronectin; post translational hydroxylation.			
XX	OS	Synthetic.	RESULT 7	AAR1702
OS	OS	Homo sapiens.	ID	AAR1702 standard; protein: 1366 AA.
XX	PN	EP992586-A2.	XX	AAR1702;
XX	PD	12-APR-2000.	XX	17-OCT-1995 (first entry)
XX	PR	07-OCT-1999; 99EP-0119184.	DE	Collagen alpha 2 (I) chain precursor.
XX	PR	09-OCT-1998; 98US-0169768.	XX	Collagen; antibody; immunoassay; metabolism; diagnosis; monitoring; disorder; osteoporosis; metastatic progression; Paget's disease; hyperthyroidism; bone; resorption; rheumatoid arthritis; osteoarthritis; vasculitis syndrome.
XX	PA	(USSU ) US SURGICAL CORP.	XX	Homo sapiens.
XX	PI	Gruskin EA, Buechler DD, Zhang G, Connolly K;	PN	WO9508115-A.
XX	DR	WPI; 2000-259138/23.	XX	
DR	N-PSDB; AAA12511.	PD	23-MAR-1995.	
XX	PT	Production of extracellular matrix proteins containing 4-trans-hydroxyproline results in native self aggregating proteins, useful on medical implants -	XX	
PT	Example 14; FIG 50A-E; 260pp; English.	PR	19-SEP-1994; 94WO-DK003148.	
XX	PS	The specification describes a method for producing an extracellular matrix protein or its fragment. The extracellular matrix protein is capable of self aggregating in a cell which does not ordinarily hydroxylate prolines. The method comprises optimising a nucleic acid sequence for expression in the cell by substitution of codons preferred by that cell for naturally occurring codons not preferred by the cell; incorporating the nucleic acid sequence into the cell; and contacting the cell with a hypertonic growth medium containing at least one amino acid, selected from the group consisting of trans-4-hydroxyproline and 3-hydroxyproline to allow at least one of the amino acids to be assimilated into the cell and incorporated into the extracellular matrix protein. The method may be used to make host cells assimilate and incorporate trans-4-hydroxyproline into proteins. This is especially useful in the recombinant production of proteins such as collagen, fibrinogen and fibronectin whose ability to self aggregate and produce functional proteins depends on the post translational hydroxylation of proline. The method is also useful in studying the structure and function of polypeptides which do not normally contain trans-4-hydroxyproline. The present sequence represents a human collagen 1 ( $\alpha_2$ ) helical	PA	17-SEP-1993; 93DK-0001040.
XX	PS	Disclosure (Appendix A); Page 51; 8pp; English.	PA	(OSTB-) OSTEOEMETER AS.
XX	PS	Determination of collagen fragments in body fluid can be achieved by immunoassay using antibodies directed against synthetic peptides derived from collagen which contain sites of potential crosslinking. The method is used to diagnose and monitor treatment of disorders of collagen metabolism (degradation of type I collagen may indicate osteoporosis, metastatic progression, Paget's disease, hyperthyroidism, metabolic syndrome, or other conditions involving excessive bone resorption; degradation of type II collagen may indicate rheumatoid arthritis or osteoarthritis; and of type III collagen, vasculitis (syndrome). The method can be used to assess the toxicity of a compound and to test drugs for their effect on collagen metabolism.	PI	Bonde M, Qvist P;
XX	PS	Sequence 1366 AA;	XX	WPI; 1995-131456/17.
CC	Query Match	68.5%; Score 63; DB 16; Length 1366;	Query Match	68.5%; Score 63; DB 16; Length 1366;
CC	Best Local Similarity	73.3%; Pred. No. 1.7;	Matches	11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
CC	Matches	11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;	Qy	1 PAGPWGPNGDKGKG 15
CC	Db	1055 PAGPSGPAGKDGRIG 1069	Db	1055 PAGPSGPAGKDGRIG 1069

**RESULT 8**  
 AAY96123  
 ID AAY96123 standard; Peptide: 1366 AA.  
 XX  
 AC AAY96123;  
 XX DT 19-DEC-2000 (first entry)  
 XX DE Collagen type I alpha-2.  
 XX KW Collagen type I; osteoporosis; bone resportion; Paget's disease; hyperparathyroidism; metastasis; assay; diagnosis.  
 XX OS Homo sapiens.  
 XX PN US6110689-A.  
 XX PD 29-AUG-2000.  
 XX PR 04-NOV-1997; 97US-0963825.  
 XX PR 21-JAN-1994; 94US-0187319.  
 XX PA (OSTE-) OSTROMETER AS.  
 XX PT Bonde M, Qvist P;  
 XX DR WPI; 2000-586349/55.  
 XX Assaying type I collagen fragments for diagnosing osteoporosis in postmenopausal woman, involves contacting body fluid with synthetic collagen peptide and antibody and quantifying by competitive binding assay -  
 XX PS disclosure; Column 31-37; 41pp; English.  
 XX The present sequence is that of human type I collagen alpha-2.  
 CC The invention is based on the discovery of the presence of particular collagen fragments in body fluids of patients compared with those of healthy subjects. These fragments are generated upon collagen degradation and are partly characterised by the presence of potential sites for crosslinking. A method for assaying collagen fragments in a body fluid sample is based on the competitive binding to immunological binding partners of collagen fragments in the sample and of synthetic peptides derived from collagen and containing crosslinkable sites (see AAY96105-11). When considering the degradation of type I collagen, the assay can be used as a means of identifying excessive bone resportion, indicating the presence of osteoporosis or the metastatic progres of a malignancy. Other conditions characterized by excessive bone resportion include Paget's disease and hyperparathyroidism.

**SQ Sequence 1366 AA:**

Query Match	Score	DB	Length
Best Local Similarity	68.5%	21	1366
Matches	73.3%	Pred. No.	1.7
11; Conservative	1	Mismatches	3
Indels	0	Indels	0
Gaps	0	Gaps	0

**RESULT 9**  
 AAY56801  
 ID AAY56801 standard; protein: 1366 AA.  
 XX AC AAY56801;  
 XX DT 27-MAR-2000 (first entry)  
 XX XX Human preproalpha 2 (I) collagen.  
 XX DE Fibrillar collagen; C propeptide; SSAD; telopeptide; gelatin; sequence selection and alignment domain; prosthetic implant; foodstuff; medicine; type I collagen; human.  
 XX OS Homo sapiens.  
 XX PN EB967226-A2.  
 XX PD 29-DEC-1999.  
 XX PF 04-MAY-1999; 99EP-0303470.  
 XX PR 08-MAY-1998; 98US-0284828.  
 XX PR 10-APR-1999; 99US-0289578.  
 XX PA (COHE-) COHESIN TECHNOLOGIES INC.  
 XX Oiseen DR, Hitzeman RA, Chisholm GE;  
 XX DR WPI; 2000-074666/07.  
 XX PT New method for production of fibrillar collagen, useful for preparing telopeptide collagen fibrils and gelatin -  
 XX PS Example 1; Fig 4A-B; 30pp; English.  
 XX CC The invention provides a method for the production of fibrillar collagen. The method comprises: (a) culturing a recombinant host cell comprising a DNA encoding a fibrillar collagen monomer lacking a C propeptide SSAD (sequence selection and alignment domain); and (b) producing the fibrillar collagen. The methods are used to produce fibrillar collagen, from which telopeptide collagen fibrils can be derived. Host cells, comprising DNA encoding a collagen monomer lacking SSAD or N propeptide is used to produce gelatin. Collagen is used in biological research as a substrate for in vitro cell culture and as a component of biocompatible materials for use in prosthetic implants, sustained drug release matrices, artificial skin and wound dressing and healing devices. Gelatin is particularly useful for foodstuffs and medicine, for coating tablets and making capsules. The methods, comprising the use of collagen monomers lacking the N and/or C propeptides, result in a large increase in the production of type I collagen. The present sequence represents the human preproalpha 2 (I) collagen (GenBank Accn no: 274616).  
 CC SQ Sequence 1366 AA:  

Query Match	Score	DB	Length
Best Local Similarity	68.5%	21	1366
Matches	73.3%	Pred. No.	1.7
11; Conservative	1	Mismatches	3
Indels	0	Indels	0
Gaps	0	Gaps	0

**RESULT 10**  
 ABB5023  
 ID ABB5023 standard; Protein: 1366 AA.  
 XX AC ABB5023;  
 XX AB5023;  
 XX DT 08-FEB-2002 (first entry)  
 XX DE Collagen type I alpha-2 ovarian tumour marker protein, SEQ ID NO:76.  
 XX KW Ovarian tumour marker gene; human; overexpression; upregulation; epithelial tumour; cancer; diagnosis; prognosis; disease monitoring; identification; serous cystadenoma; borderline serous tumour; serous cystadenocarcinoma; mucinous cystadenocarcinoma; mucinous adenocarcinoma; borderline mucinous tumour; endometrioid carcinoma; undifferentiated carcinoma; clear cell adenocarcinoma; cystadenofibroma; adenofibroma; Brenner tumour; serial analysis of gene expression; SAGE;

KW	immune response pathway; cell proliferation regulation; protein folding;	DT	10-AUG-2001 (first entry)
KW	membrane localised; secreted; therapeutic target; cytostatic;	XX	
gene therapy; vaccine.		DE	Porcine alpha2(I) collagen.
XX		XX	
OS	Homo sapiens.	KW	Porcine; alpha2(I) collagen; gelatin; cytostatic; viral infection;
XX		KW	pharmaceutical; food industry; cosmetic; autoimmune disorder; vaccine;
PN	WO200175177-A2.	KW	medical; arterial sealant; bone graft; dermal implant; haemostat; cancer;
XX		KW	rheumatoid arthritis; beverage; photographic application.
PD	11-OCT-2001.	OS	Sus scrofa.
XX		XX	
PP	03-APR-2001; 2001WO-US10947.	PN	WO200134647-A2.
XX		XX	
PR	03-APR-2000; 2000US-194336P.	PD	17-MAY-2001.
XX		XX	
PA	(USSH ) US DEPT HEALTH & HUMAN SERVICES.	PP	10-NOV-2000; 2000WO-US30792.
XX		XX	
PI	Morin PJ, Sherman-Baust CA, Pizer ES, Hough CD;	PR	12-NOV-1999; 99US-0439058.
XX		PR	10-NOV-2000; 2000US-0439058.
DR	WPI: 2001-626450/72.	PA	
XX	N-PSDB; ABA83119.	XX	
PS	Claim 23: Page 119-122; 140pp; English.	XX	
XX		XX	
CC	The invention relates to methods for diagnosing and prognosis ovarian tumours in an individual via the detection and measurement of the expression of ovarian tumour marker genes (ABA83121-ABA83122, ABA83180', CC ABA83182 and ABA83184) or segments thereof (ABA83123-ABA83169, ABA83179, ABA83181 and ABA83183). The methods of the invention are useful for detecting an ovarian tumour in a patient, for identifying an individual at increased risk for developing ovarian cancer, in prognostic tests for assessing the relative severity of ovarian cancer, in tests for monitoring a patient in remission from ovarian cancer and in tests for monitoring disease status in a patient being treated for ovarian cancer. The methods can additionally be used to identify a particular tumour as being an ovarian tumour (i.e., an epithelial ovarian tumour selected from serous cystadenoma, borderline serous tumour, serous cystadenocarcinoma, mucinous cystadenoma, borderline mucinous tumour, mucinous cystadenocarcinoma, endometrioid carcinoma, undifferentiated carcinoma, clear cell adenocarcinoma, cystadenofibroma, adenofibroma and Brenner tumour. The ovarian tumour marker genes of the invention were identified using SAGE (serial analysis of gene expression) and were found to be overexpressed in a broad variety of ovarian epithelial tumour cells relative to normal ovarian epithelial cells. The marker genes are implicated in immune response pathways, in the regulation of cell proliferation and in protein folding, and many of these are membrane-localised or secreted. In addition to their use as diagnostic and prognostic markers, the ovarian tumour marker genes or their encoded proteins may be used as therapeutic targets for the treatment and prevention of ovarian cancer. Sequences ABB50257-ABB50299 represent proteins encoded by ovarian tumour marker genes of the invention.	XX	
XX	Sequence 1366 AA:	XX	
SQ	Query Match 68.5%; Score 63; DB 22; Length 1366; Best Local Similarity 73.3%; Pred. No. 1.7; Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;	XX	
Qy	1 PAGPGPGPKDGKVG 15	Qy	1 PAGPGPGPKDGKVG 15
Db	1055 PAGPGGPKDGKVG 1069	Db	1055 PAGPGGPKDGKVG 1069
XX	RESULT 12	XX	RESULT 12
RESULT 11		XX	AAR59751
AAE2536		XX	AAR59751 standard; protein; 1418 AA.
ID	AAE02536 standard; Protein; 1366 AA.	AC	AAR59751;
XX		XX	14-FEB-1995 (first entry)
AC	AAE2536;	XX	Type II collagen.
XX		DE	

KW Collagen; triple helix; articular cartilage; collagenase;  
 KW degradation; monoclonal antibody; epitope; matrix;  
 KW metalloproteinase.  
 OS Homo sapiens.  
 XX  
 PH Key Location/Qualifiers  
 PT 1..24  
 FT /label= Signal peptide.  
 XX  
 PN WO9414070-A.  
 XX  
 PD 23-JUN-1994.  
 XX  
 PR 04-DEC-1992; 92US-0984123.  
 PR 17-JUL-1995; 95US-0448501.  
 XX  
 PA (SHRI-) SHRINERS HOSPITALS FOR CHILDREN.  
 PI Billinghurst RC, Poole AR,  
 XX DR WPI; 2001-006136/01.  
 XX  
 PT Detecting cartilage degradation useful for early detection of arthritis  
 PT or joint damage by contacting the biological sample with an antibody  
 PT that binds to an epitope of unwound type II collagen chains but not to  
 PT a native helical collagen -  
 XX  
 PR Example 1; Fig 1; 58pp; English.  
 XX  
 PA (SHRI-) SHRINERS HOSPITALS FOR CRIPPLED CHILDREN.  
 PI Hollander AP, Poole AR,  
 XX DR WPI; 1994-234222/28.  
 XX  
 PT Determinn. of cartilage degradation - using a monoclonal antibody  
 PT to measure the amt. of unwound collagen or fragments in samples  
 XX  
 PS Disclosure; Figure 1; 119pp; English.  
 XX  
 CC Type II collagen constitutes the bulk of the fibrillar backbone of  
 CC which is composed of a tightly wound triple helix  
 CC that can only be cleaved by the metalloproteinase collagenase to  
 CC produce 3/4 and 1/4 length alpha chain fragments. The destruction of  
 CC articular cartilage is due, in part, to the degradation of collagen.  
 CC Incapable of maintaining its helical structure at physiological  
 CC temperatures, collagenase-cleaved collagens unwind and become  
 CC susceptible to further degradation by other proteinases. By  
 CC producing monoclonal antibodies directed against epitopes which are  
 CC only revealed when collagen is unwound, the antibodies provide a  
 CC means of determining the degradation of cartilage in a biological  
 CC sample. The antibodies do not bind to native helical collagen.  
 CC Epitopes used in the production of such antibodies are described in  
 CC AAR59749, AAR59750 and AAR67742.  
 XX  
 SQ Sequence 1418 AA:  
 Query Match 64.1%; Score 59; DB 22; Length 1418;  
 Best Local Similarity 66.7%; Pred. No. 6.3.;  
 Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 PAGPGPGNGKDGKV 15  
 QY 1 ||| ||:|||| :|  
 Db 1096 PPGPVGPGSKDGAMG 1110  
 XX  
 RESULT 14  
 AAR9480  
 ID AAR9480 standard; Protein; 1442 AA.  
 XX  
 AC AAR9480;  
 XX DT 19-JAN-1996 (first entry)  
 XX DE Rat type II collagen.  
 XX KW Collagen; bone progenitor; gene transfer; gene therapy; osteoporosis;  
 KW osteotomy; bone repair; osteotropic; *Pichia pastoris*.  
 XX OS Rattus sp.  
 XX PN WO9522611-A2.  
 XX  
 PD 24-AUG-1995.  
 XX  
 PR 21-FEB-1995; 95WO-US02251.  
 XX  
 PR 30-SEP-1994; 94US-0316650.  
 PR 18-FEB-1994; 94US-0199780.  
 XX  
 PA (UNMI ) UNIV MICHIGAN.  
 XX DR WPI; 1995-302717/39.  
 XX  
 PT Transferring nucleic acid into bone progenitor cell(s) - using a  
 PT bone compatible matrix, for treatment of fracture(s) and  
 PT osteoporosis.  
 XX  
 PS Disclosure; Page 197-208; 317pp; English.  
 XX  
 CC Human, rat and mouse collagen type II (given in AAR79479-81).

CC respectively) can be used to stimulate bone progenitor cells as a means of treating bone-related diseases in association with an osteotropic gene.

CC  
CC  
XX  
SQ Sequence 1442 AA;

Query Match 62.0%; Score 57; DB 16; Length 1442;  
Best Local Similarity 73.3%; Pred. No. 12; Mismatches 11;  
Matches 1; Conservative 0; Mismatches 4; Indels 4; Gaps 0;

Qy 1 PAGPWGPNGKDGKVG 15  
Db 871 PAGPPGPAGSKDGPKG 885

RESULT 15  
AAB4187  
ID AAB4187 standard; Protein; 97 AA.  
AC AAB4187;  
XX DT 08-FEB-2001 (first entry)

DE Human cancer associated protein sequence SEQ ID NO:1632.

XX Human; cancer associated gene; cancer antigen; detection; cancer; diagnosis; cyostatic; proliferative; vulneary; immunomodulator; antidiabetic; antiasthmatic; antirheumatic; antiarthritic; antiviral; antiinflammatory; antithyroid; antibacterial; cardivral; dermatological; neuroprotective; thrombolytic; coagulant; nootropic; vasoconstrictor; antipsoriatic; antiangiogenic; gene therapy; inflammation; immune disorder; haematopoietic cell disorder; autoimmune disorder; allergic reaction; graft versus host disease; organ rejection; haemostatic; thrombolytic; cardiovascular disorder; infection; neurological disease; drug screening.

Homo sapiens.

XX WO200055350-A1.

XX PR 12-MAR-1999; 99US-0124270.

XX PD 21-SEP-2000.

XX PF 08-MAR-2000; 2000WO-US05882.

XX PR 12-MAR-1999; 99US-0124270.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Rosen CA; Roben SM;  
XX DR WPI; 2000-587533/55.  
XX N-PSDB; AAC78396.

XX Novel isolated nucleic acids comprising sequences encoding peptides useful for treating or diagnosing e.g. cancer -  
XX PT Claim 11; Page 2302; 2352pp; English.

XX AAC77607 to AAC78448 encode the human cancer associated proteins given in AAC77607 to AAC78448. The proteins can have activities based on the tissues and cells the genes are expressed in. Example of activities include: cytostatic; proliferative; vulneary; immunomodulator; antidiabetic; antiasthmatic; antirheumatic; antiarthritic; antiinflammatory; antithyroid; antibacterial; antiviral; dermatological; neuroprotective; cardivral; thrombolytic; coagulant; nootropic; vasoconstrictor; antipsoriatic and antiangiogenic. The polynucleotides and polypeptides can be used for preventing, treating or ameliorating medical conditions and diagnosing pathological conditions. Polynucleotides, polypeptides, antibodies, agonists and antagonists from the present invention may be used to treat immune disorders by activating or inhibiting the proliferation, differentiation or mobilisation of immune cells, to treat disorders of haematopoietic cells, autoimmune disorders, allergic reactions, graft versus host disease and organ

CC rejection, modulate haemostatic or thrombolytic activity, modulate inflammation, cancers, cardiovascular disorders, neurological disease and bacterial or viral infections. The peptides, nucleotides, antibodies, agonists and antagonists may be also be used in drug screens. AAC78449 to AAC78457 and AAB44240 represent sequences used in the exemplification of the present Invention.

CC  
CC  
XX SQ Sequence 97 AA;

Query Match 60.9%; Score 56; DB 21; Length 97;  
Best Local Similarity 66.7%; Pred. No. 1.1; Mismatches 10;  
Matches 1; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 PAGPWGPNGKDGKVG 15  
Db 13 PVGPXGPCKDGTXG 27

RESULT 16  
AAR79481  
ID AAR79481 standard; Protein; 54 AA.  
AC AAR79481;  
XX DT 19-JAN-1996 (first entry)

XX DE Mouse type II collagen.

XX KW Collagen; bone progenitor; gene transfer; gene therapy; osteoporosis; osteotomy; bone repair; osteotropic; Fichia pastoris.

XX OS Mus sp.

XX PN WO9522611-A2.

XX PD 24-AUG-1995.

XX PR 21-FEB-1995; 95WO-US02251.

XX PR 30-SBP-1994; 94US-0316650.

XX PR 18-FEB-1994; 94US-0199780.

XX PA (UNM-) UNIV MICHIGAN.

XX DR WPI; 1995-302717/39.

XX PT Transferring nucleic acid into bone progenitor cell(s) - using a bone compatible matrix, for treatment of fracture(s) and osteoporosis.

XX PS Disclosure; Page 210; 317pp; English.

XX CC Human, rat and mouse collagen type II (given in AAR79479-81, respectively) can be used to stimulate bone progenitor cells as a means of treating bone-related diseases in association with an osteotropic gene.

XX SQ Sequence 54 AA;

Query Match 59.8%; Score 55; DB 16; Length 54;  
Best Local Similarity 66.7%; Pred. No. 0.87; Mismatches 10;  
Matches 1; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 1 PAGPWGPNGKDGKVG 15  
Db 20 PPGVGVPSGKDGANG 34

RESULT 17  
AAR71703  
ID AAR71703 standard; protein; 1418 AA.

XX AC AAR71703;

XX	17-OCT-1995	(first entry)
DT	DE	Collagen alpha 1 (II) chain precursor.
XX	XX	
KW	Collagen; antibody; immunoassay; metabolism; diagnosis; monitoring;	
KW	disorder; osteoporosis; metastatic progression; Paget's disease;	
KW	hyperthyroidism; bone; resorption; rheumatoid arthritis;	
KW	osteoarthritis; vasculitis syndrome.	
OS	Homo sapiens.	
XX		
PN	WO9508115-A.	
XX		
PD	23-MAR-1995.	
XX		
PT	19-SEP-1994; 94WO-DK00348.	
XX		
PR	17-SEP-1993; 93DK-0001040.	
XX		
PI	Bonde M, Qvist P;	
XX		
DR	WPI: 1995-131456/17.	
XX		
PT	Assaying collagen fragments in body fluid by immunoassay - using postmenopausal woman, involves contacting body fluid with synthetic collagen Peptide and antibody and quantifying by competitive binding assay.	
XX		
PA	(OSTE-) OSTEOmeter AS.	
XX		
PT	Bonde M, Qvist P;	
XX		
PS	WPI: 2000-586349/55.	
XX		
PT	Assaying type I collagen fragments for diagnosing osteoporosis in postmenopausal woman, involves contacting body fluid with synthetic collagen Peptide and antibody and quantifying by competitive binding assay.	
XX		
PT	Disclosure; Column 37-46; 41pp; English.	
XX		
CC	Determination of collagen fragments in body fluids can be achieved by immunoassay using antibodies directed against synthetic peptides derived from collagen which contain sites of potential crosslinking.	
CC	The method is used to diagnose and monitor treatment of disorders of collagen metabolism (degradation of type I collagen may indicate osteoporosis), metastatic progression, Paget's disease, hyperthyroidism or other conditions involving excessive bone resorption; degredation of type II collagen may indicate rheumatoid arthritis or osteoarthritis; and of type III collagen, vacuilitis syndrome). The method can also be used to assess the toxicity of a compound and to test drugs for their effect on collagen metabolism.	
CC	Sequence 1418 AA:	
XX		
SQ	Query Match 59.8%; Score 55; DB 16; Length 1418; Best Local Similarity 66.7%; Pred. No. 23; Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;	
CC	Sequence 1418 AA:	
XX		
SQ	Query Match 59.8%; Score 55; DB 21; Length 1418; Best Local Similarity 66.7%; Pred. No. 23; Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;	
CC	Sequence 1418 AA:	
XX		
RESULT 19		
AAW61562		
ID AAW61562 standard; protein; 1487 AA.		
XX		
AC AAW61562;		
XX		
DT 02-NOV-1998 (first entry)		
XX		
DE Human type II collagen alpha-chain protein.		
XX		
DB 1096 PPGFVGPSKDGAN 1110		
XX		
RESULT 19		
AAW61562		
ID AAW61562 standard; protein; 1487 AA.		
XX		
AC AAW61562;		
XX		
DT 02-NOV-1998 (first entry)		
XX		
DE Human type II collagen alpha-chain protein.		
XX		
DB 1096 PPGFVGPSKDGAN 1110		
XX		
RESULT 19		
AAW61562		
ID AAY6124 standard; peptide; 1418 AA.		
XX		
AC AAY6124;		
XX		
DT 19-DEC-2000 (first entry)		
XX		
DE Collagen type II alpha-1.		
XX		
KW Collagen type II; rheumatoid arthritis; osteoarthritis; assay; diagnosis.		
XX		
KW diagnosis.		
OS Homo sapiens.		
XX		
PN US6110689-A.		
XX		

DR WPI; 1998-447376/38.

XX Immunoassay kit containing two antibodies recognising coupled PT epitope(s) on collagen fragments - and new antibodies, for PT diagnosing arthritis etc., also prognosis and screening for PT anti-arthritis agents or inhibitors of matrix metallo-protease XX Disclosure; Fig 2; 57pp; English.

This sequence represents the human type II collagen alpha-chain which is used in a method to produce an immunoassay kit comprising of two antibodies (Ab1 and Ab2), mono- or poly-clonal, or their fragments, that bind to two C-telopeptide coupled epitopes (C-telopeptide indicates any type II collagen fragment that is released from degraded cartilage). The kits are designed for sandwich immunoassays, specifically enzyme-linked immunosorbent assay (ELISA), and C-telopeptide is systemic (present in urine, serum or synovial fluid). The kits are used for therapy, diagnosis (e.g. routine screening for arthritis and other cartilage diseases, also to diagnose growth disorders), prognosis (e.g. monitoring progression of rheumatoid arthritis and osteoarthritis, or monitoring treatment with growth hormone) and for drug screening (to identify, and assess efficacy of, anti-arthritis agents and matrix metalloprotease inhibitors). C-telopeptide, derived from the N-terminus of the alpha 1 chain, have increased resistance to proteolysis, so can accumulate in vivo to a concentration that allows accurate measurement by immunoassay.

XX Sequence 1487 AA;

SQ Query Match 59.8%; Score 55; DB 19; Length 1487;

Best Local Similarity 66.7%; Pred. No. 24; Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 1 PAGPWGPNGKDGKVG 15  
Db 1165 PPGPVGPGSKDGANG 1179

RESULT 20

ID AAR71704 standard; protein; 1078 AA.  
XX AAR71704;  
AC AAR71704;

XX DT 17-Oct-1995 (first entry)  
XX DE Collagen alpha 1 (III) chain precursor.  
XX KW Collagen; antibody; immunoassay; metabolism; diagnosis; disorder; osteoporosis; metastatic protraction; Paget's disease; hyperthyroidism; bone; resorption; rheumatoid arthritis; osteoarthritis; vasculitis syndrome.

OS Homo sapiens.

XX PN US6110689-A.  
XX PD 29-AUG-2000.  
XX PR 04-NOV-1997; 97US-0963825.  
XX PR 21-JAN-1994; 94US-0187319.

XX PA (OSTE-) OSTEOMETER AS.  
XX PI Bonde M, Qvist P;  
XX DR WPI; 2000-586349/55.

XX PT Assaying type I collagen fragments for diagnosing osteoporosis in postmenopausal woman, involves contacting body fluid with synthetic collagen peptide and antibody and quantifying by competitive binding assay

XX Disclosure; Column 46-51; 41pp; English.

XX PR 17-SEP-1993; 93DK-0001040.  
XX PA (OSTE-) OSTEOMETER AS.

XX PI Bonde M, Qvist P;  
XX DR WPI; 1995-131456/17.

XX Assaying collagen fragments in body fluid by immunoassay - using antibodies raised against synthetic peptide(s) contg. potential crosslinking sites, to diagnose and monitor disorders of collagen metabolism, e.g. osteoporosis.

PS Disclosure (Appendix A); Page 55; 87pp; English.

XX CC Determination of collagen fragments in body fluids can be achieved by immunoassay using antibodies directed against synthetic peptides derived from collagen which contain sites of potential crosslinking. The method is used to diagnose and monitor treatment of disorders of collagen metabolism (degradation of type I collagen may indicate osteoporosis, metastatic progression, Paget's disease, hyperthyroidism or other conditions involving excessive bone resorption; degradation of type II collagen may indicate rheumatoid arthritis or osteoarthritis; and of type III collagen, vasculitis syndrome). The method can also be used to assess the toxicity of a compound and to test drugs for their effect on collagen metabolism.

XX Sequence 1078 AA;

Query Match 58.7%; Score 54; DB 16; Length 1078;

Best Local Similarity 66.7%; Pred. No. 24; Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 PAGPWGPNGKDGKVG 15  
Db 994 PVGPGSPPKGKGTSG 1008

RESULT 21

ID AAY96125 standard; Peptide; 1078 AA.

XX AC AAY96125;

XX DT 19-DEC-2000 (first entry)

XX DE Collagen type III alpha-1.

XX KW Collagen type III; vasculitis syndrome; assay; diagnosis.

XX OS Homo sapiens.

XX PN US6110689-A.

XX PR 04-NOV-1997; 97US-0963825.

XX PR 21-JAN-1994; 94US-0187319.

XX PA (OSTE-) OSTEOMETER AS.

XX PI Bonde M, Qvist P;

XX DR WPI; 2000-586349/55.

XX PT Assaying type I collagen fragments for diagnosing osteoporosis in postmenopausal woman, involves contacting body fluid with synthetic collagen peptide and antibody and quantifying by competitive binding assay

XX Disclosure; Column 46-51; 41pp; English.

XX The present sequence is that of human type III collagen alpha-1.

CC The invention is based on the discovery of the presence of particular collagen fragments in body fluids of patients compared with those of healthy subjects. These fragments are generated upon collagen degradation and are partly characterised by the presence of potential sites for crosslinking. A method for assaying collagen fragments in a body fluid sample is based on the competitive binding to immunological binding partners of collagen fragments in the sample and of synthetic peptides derived from collagen and containing crosslinkable sites (see AAY96118-21). When considering the degradation of type III collagen, the assay can be used as a means of identifying the presence of vasculitis syndrome.



FT Peptide 586..591  
 FT /label= GXYGX'Y'\_motif  
 FT 592..597  
 FT Peptide /label= GXYGX'Y'\_motif  
 FT 598..603  
 FT Peptide /label= GXYGX'Y'\_motif  
 FT 604..609  
 FT /label= GXYGX'Y'\_motif  
 FT 610..615  
 FT Peptide /label= GXYGX'Y'\_motif  
 FT 616..621  
 FT Peptide /label= GXYGX'Y'\_motif  
 FT 622..627  
 FT /label= GXYGX'Y'\_motif  
 FT 628..633  
 FT Peptide /label= GXYGX'Y'\_motif  
 FT 634..639  
 FT Peptide /label= GXYGX'Y'\_motif  
 FT 640..665  
 FT /label= GXYGX'Y'\_motif  
 FT 657..662  
 FT Peptide /label= GXYGX'Y'\_motif  
 FT 677..682  
 FT Peptide /label= GXYGX'Y'\_motif  
 FT 683..688  
 FT /label= GXYGX'Y'\_motif  
 FT 689..694  
 FT Peptide /label= GXYGX'Y'\_motif  
 FT 695..700  
 FT Peptide /label= GXYGX'Y'\_motif  
 FT 707..712  
 FT /label= GXYGX'Y'\_motif  
 FT 713..718  
 FT Peptide /label= GXYGX'Y'\_motif  
 FT 735..740  
 FT Peptide /label= GXYGX'Y'\_motif  
 FT 741..746  
 FT /label= GXYGX'Y'\_motif  
 FT 747..752  
 FT Peptide /label= GXYGX'Y'\_motif  
 FT 759..764  
 FT Peptide /label= GXYGX'Y'\_motif  
 FT 765..770  
 FT Peptide /label= GXYGX'Y'\_motif  
 FT 771..776  
 FT Peptide /label= GXYGX'Y'\_motif  
 FT 787..792  
 FT Peptide /label= GXYGX'Y'\_motif  
 FT 793..798  
 FT Peptide /label= GXYGX'Y'\_motif  
 FT 799..804  
 FT Peptide /label= GXYGX'Y'\_motif  
 FT 815..820  
 FT Peptide /label= GXYGX'Y'\_motif  
 FT 821..826  
 FT Peptide /label= GXYGX'Y'\_motif  
 FT 827..832  
 FT Peptide /label= GXYGX'Y'\_motif  
 FT 833..838  
 FT Peptide /label= GXYGX'Y'\_motif  
 FT 839..844  
 FT Peptide /label= GXYGX'Y'\_motif  
 FT 845..850  
 FT Peptide /label= GXYGX'Y'\_motif  
 FT 863..868  
 FT Peptide /label= GXYGX'Y'\_motif  
 FT 869..874  
 FT Peptide /label= GXYGX'Y'\_motif  
 FT 875..880  
 FT Peptide /label= GXYGX'Y'\_motif  
 FT 891..896  
 FT Peptide /label= GXYGX'Y'\_motif  
 FT 897..902  
 FT Peptide

FT Peptide /label= GXYGX'Y'\_motif  
 FT 903..908  
 FT Peptide /label= GXYGX'Y'\_motif  
 FT 911..916  
 FT Peptide /label= GXYGX'Y'\_motif  
 FT 917..922  
 FT Peptide /label= GXYGX'Y'\_motif  
 FT 928..933  
 FT Peptide /label= GXYGX'Y'\_motif  
 FT 934..939  
 FT Peptide /label= GXYGX'Y'\_motif  
 FT 956..961  
 FT Peptide /label= GXYGX'Y'\_motif  
 FT 962..967  
 FT Peptide /label= GXYGX'Y'\_motif  
 FT 968..973  
 FT Peptide /label= GXYGX'Y'\_motif  
 FT 1126..1131  
 FT Peptide /label= GXYGX'Y'\_motif  
 FT 1145..1150  
 FT Peptide /label= GXYGX'Y'\_motif  
 FT 1193..1198  
 FT Peptide /label= GXYGX'Y'\_motif

XX PN US5643783-A.  
 XX PD 01-JUL-1997.  
 XX PR 01-DEC-1993;  
 XX PR 01-DEC-1993; 93US-0159784.  
 XX PR 01-DEC-1993; 93US-0159784.  
 XX PA (HARD ) HARVARD COLLEGE.  
 XX PI Oh SP , Olsen BR;  
 XX DR WPI: 1997-3502A7/32.  
 XX DR N-PSB; AAT84485.  
 XX PT Disclosure; Fig 2; 35pp; English.  
 Query Match 58.7%; Score 54; DB 18; Length 1288;  
 Best Local Similarity 66.7%; Pred No. 29;  
 Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 PT degeneration  
 XX PS Nucleic acid encoding human alpha-1 collagen - for production of recombinant alpha-1 collagen, for use in the treatment of cartilage degeneration

RESULT 24  
 ID AAW92297 standard; peptide; 1288 AA.  
 ID AAW92297  
 ID AAW92297;  
 AC AC  
 XX DT 28-APR-1999 (first entry)  
 XX DE Mouse alpha-1 (XVIII) collagen chain common sequence M018(common)28.  
 XX KW Human; type XVIII collagen; liver disease; cirrhosis; detection; hepatocellular carcinoma; diagnosis.  
 XX OS Mus sp.  
 XX PN WO9856399-A1.  
 XX PD 17-DEC-1998.  
 XX

PF 12-JUN-1998; 98HQ-US12327.  
 XX PR 03-APR-2000; 2000US-194336P.  
 XX  
 XX PR (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
 PA XX Morin PJ, Sherman-Baust CA, Pizer ES, Hough CD;  
 PA PT DR WPI; 2001-626450/72.  
 PA XX N-PSDB; ABA83117.  
 (FIFT-) ACAD FINLAND.  
 PA INST NAT SANTE & RECH MEDICALE.  
 PA (INRM )  
 XX PT Clement B, Pihla-Janiemi T, Rehn M;  
 XX DR WPI; 1999-070292/06.  
 PT Diagnosis and monitoring of liver disease by measuring collagen type  
 PT XVIII levels - with elevated levels indicative of disease,  
 especially cirrhosis or hepatocellular carcinoma  
 XX PS Example 6; FIG 8; 56pp; English.  
 CC A method has been developed for the detecting liver disease. The method  
 comprises: (a) reacting a patient sample with antibodies (Ab) specific  
 for collagen type XVIII (Col18); (b) measuring the amount of Ab-antigen  
 complex (C) formed as indicator of the mount of Col18 present; (c)  
 similar analysis of a non-diseased control; and (d) comparing the  
 amounts of Col18 in the two samples to detect presence or progression of  
 disease. Elevated levels of Col18 are: (i) indicative of disease,  
 specifically cirrosis; and (ii) predictive of the prognosis of disease,  
 specifically hepatocellular carcinoma (there is a relationship between  
 Col18 mRNA levels and tumour size and necrosis, and survival times are  
 significantly higher in patients with higher Col18 levels). The method  
 provides non-invasive, early and accurate diagnosis of liver disease.  
 CC The present sequence represents the sequence common to mouse alpha-1  
 CC (XVIII) collagen chain from the present invention.  
 CC Sequence 1288 AA:  
 Query Match 58.7%; Score 54; DB 20; Length 1288;  
 Best local Similarity 66.7%; Pred. No. 29;  
 Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 QY 1 PAGPWGPNGKDGKVG 15  
 DB 341 PPGPQGPQGPDKGPG 355  
 SQ RESULT 25  
 ABB50291 ID ABB50291 standard; Protein; 1466 AA.  
 AC XX  
 AC XX  
 AC XX  
 DT 08-FEB-2002 (first entry)  
 DE Collagen type III alpha-1 ovarian tumour marker protein, SEQ ID NO:72.  
 XX Ovarian tumour marker gene; human; overexpression; upregulation;  
 KW epithelial tumour; cancer; diagnosis; disease monitoring;  
 KW identification; serous cystadenoma; borderline serous tumour;  
 KW serous cystadenocarcinoma; mucinous cystadenocarcinoma;  
 KW mucinous cystadenoma; borderline mucinous tumour; endometrioid carcinoma;  
 KW undifferentiated carcinoma; clear cell adenocarcinoma; cystadenofibroma;  
 KW adenofibroma; Brenner tumour; serial analysis of gene expression; SAGE;  
 KW immune response pathway; cell proliferation regulation; protein folding;  
 KW membrane localised; secreted; therapeutic target; cytostatic;  
 KW gene therapy; vaccine.  
 XX Homo sapiens.  
 XX WO20017177-A2.  
 PN 11-OCT-2001.  
 PD XX  
 PF 03-APR-2001; 2001WO-US10947.  
 XX  
 PR 03-APR-2000; 2000US-194336P.  
 XX  
 XX PR (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
 PA XX Morin PJ, Sherman-Baust CA, Pizer ES, Hough CD;  
 PA PT DR WPI; 2001-626450/72.  
 PA XX N-PSDB; ABA83117.  
 PT Detecting and identifying ovarian tumor, identifying increased risk for  
 PT developing ovarian cancer, and determining effectiveness of ovarian tumor marker  
 PT gene -  
 PS Claim 23; Page 114-117; 140pp; English.  
 CC The invention relates to methods for diagnosing and prognosing ovarian  
 CC tumours in an individual via the detection and measurement of the  
 expression of ovarian tumour marker genes (ABA83081-ABA83122, ABA83180,  
 CC ABA83182 and ABA83184) or segments thereof (ABA83123-ABA83169, ABA83179,  
 CC ABA83181 and ABA83183). The methods of the invention are useful for  
 detecting an ovarian tumour in a patient, for identifying an individual  
 CC at increased risk for developing ovarian cancer, in prognostic tests for  
 assessing the relative severity of ovarian cancer, in tests for  
 CC monitoring a patient in remission from ovarian cancer and in tests for  
 monitoring disease status in a patient being treated for ovarian cancer.  
 CC The methods can additionally be used to identify a particular tumour as  
 being an ovarian tumour (i.e., an epithelial ovarian tumour selected from  
 CC serious cystadenoma, borderline serous tumour, serious cystadenocarcinoma,  
 CC mucinous cystadenoma, borderline mucinous tumour, mucinous  
 CC cystadenocarcinoma, endometrioid carcinoma, undifferentiated carcinoma,  
 CC clear cell adenocarcinoma, cystadenofibroma, adenofibroma and Brenner  
 CC tumour. The ovarian tumour marker genes of the invention were found to be  
 CC overexpressed in a broad variety of ovarian epithelial tumour genes  
 CC relative to normal ovarian epithelial cells. The marker genes are  
 CC implicated in immune response pathways, in the regulation of cell  
 proliferation and in protein folding, and many of these are membrane-  
 CC localised or secreted. In addition to their use as diagnostic and  
 prognostic markers, the ovarian tumour marker genes or their encoded  
 CC proteins may be used as therapeutic targets for the treatment and  
 prevention of ovarian cancer. Sequences ABB50257-ABB50299 represent  
 CC proteins encoded by ovarian tumour marker genes of the invention.  
 XX Sequence 1466 AA:  
 SQ RESULT 26  
 ABB50291 ID ABB50291 standard; Protein; 1466 AA.  
 AC XX  
 AC XX  
 AC XX  
 DT 10-AUG-2001 (first entry)  
 DE Bovine alpha(I,II) collagen #1.  
 XX Bovine; alpha(I,II) collagen; gelatin; cytostatic; viral infection;  
 KW pharmaceutical; food industry; cosmetic; autoimmune disorder; vaccine;  
 KW medical; arterial sealant; bone graft; dermal implant; haemostat; cancer;  
 KW rheumatoid arthritis; beverage; photographic application.  
 XX Bos sp.  
 PN WO200134647-A2.

XX  
PD 17-MAY-2001.  
XX  
PP 10-NOV-2000; 2000WO-US30792.  
XX  
PR 12-NOV-1999; 99US-0439058.  
PR 10-NOV-2000; 2000US-0439058.  
PA (FIBR-) FIBROGEN INC.  
XX  
PI Bell MP, Neff TB, Polarek JW, Seeley TW;  
XX  
DR WPI; 2001-335911/35.  
N-PSDB; AAD6574.  
XX  
PT Novel isolated and purified bovine or porcine collagens and gelatins  
PT useful in medical, pharmaceutical, food and cosmetic industries, as  
PT vaccine, and for treating autoimmune disorders, infections and cancer  
PT  
XX  
PS Example 2; Fig 4; 168pp; English.  
XX  
CC The present sequence is bovine alpha<sub>1</sub>(III) collagen. The present  
invention relates to recombinant synthesis of collagens and gelatins  
derived from animals. Collagen is useful in medical, pharmaceutical,  
food and cosmetic industries. Collagen is an important component of  
arterial sealants, bone grafts, drug delivery system, dermal implants,  
haemostats, and incontinence implants, and for treating autoimmune  
disorders such as rheumatoid arthritis. Collagen is useful in food  
products such as sausages casings, and in cosmetics or facial and skin  
formulations for treating viral infections, autoimmune diseases and  
cancer. Gelatin is useful in the manufacture or as a component of  
various pharmaceutical and medical devices and products, in food and  
beverage industries, in hair care and skin care products, as a glue or  
adhesive in various manufacturing processes, as a light-sensitive coating  
in various electronic devices, as photoresist base in photolithographic  
processes, in printing and photographic applications, in laboratory  
application, and as a component in various gels used for biochemical and  
electrophoretic analysis, including enzymographic gels.  
XX  
SQ Sequence 1466 AA;  
Query Match 58.7%; Score 54; DB 22; Length 1466;  
Best Local Similarity 66.7%; Pred. No. 33;  
Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
Qy 1 PAGPWNPGNGDGKVG 15  
Db 1141 PVGPGSGPPGKGDGASG 1155  
  
RESULT 27  
AAE02534  
ID AAE02534 standard; Protein; 1466 AA.  
AC AAE02534;  
XX  
DT 10-AUG-2001 (first entry)  
DE Bovine alpha<sub>1</sub>(III) collagen #2.  
KW Bovine; alpha<sub>1</sub>(III) collagen; gelatin; cytostatic; viral infection;  
KW pharmaceutical; food industry; cosmetic; autoimmune disorder; vaccine;  
KW medical; arterial sealant; bone graft; dermal implant; haemostat; cancer;  
KW rheumatoid arthritis; beverage; photographic application.  
XX  
OS Bos sp.  
XX  
PN WO200134647-A2.  
XX  
PD 17-MAY-2001.  
XX  
PP 10-NOV-2000; 2000WO-US30792.  
XX  
PR 12-NOV-1999; 99US-0439058.  
PR 10-NOV-2000; 2000US-0439058.  
XX  
PS Example 2; Fig 6; 168pp; English.  
XX  
CC The present sequence is bovine alpha<sub>1</sub>(III) collagen. The present  
invention relates to recombinant synthesis of collagens and gelatins  
derived from animals. Collagen is useful in medical, pharmaceutical,  
food and cosmetic industries. Collagen is an important component of  
arterial sealants, bone grafts, drug delivery system, dermal implants,  
haemostats, and incontinence implants, and for treating autoimmune  
disorders such as rheumatoid arthritis. Collagen is useful in food  
products such as sausages casings, and in cosmetics or facial and skin  
formulations for treating viral infections, autoimmune diseases and  
cancer. Gelatin is useful in the manufacture or as a component of  
various pharmaceutical and medical devices and products, in food and  
beverage industries, in hair care and skin care products, as a glue or  
adhesive in various manufacturing processes, as a light-sensitive coating  
in various electronic devices, as photoresist base in photolithographic  
processes, in printing and photographic applications, in laboratory  
application, and as a component in various gels used for biochemical and  
electrophoretic analysis, including enzymographic gels.  
XX  
SQ Sequence 1466 AA;  
Query Match 58.7%; Score 54; DB 22; Length 1466;  
Best Local Similarity 66.7%; Pred. No. 33;  
Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
Qy 1 PAGPWNPGNGDGKVG 15  
Db 1141 PVGPGSGPPGKGDGASG 1155  
  
RESULT 28  
AAE02537  
ID AAE02537 standard; Protein; 1466 AA.  
AC AAE02537;  
XX  
DT 10-AUG-2001 (first entry)  
DE Porcine alpha<sub>1</sub>(III) collagen.  
KW Porcine; alpha<sub>1</sub>(III) collagen; gelatin; cytostatic; viral infection;  
KW pharmaceutical; food industry; cosmetic; autoimmune disorder; vaccine;  
KW medical; arterial sealant; bone graft; dermal implant; haemostat; cancer;  
KW rheumatoid arthritis; beverage; photographic application.  
XX  
OS Sus scrofa.  
XX  
PN WO200134647-A2.  
XX  
PD 17-MAY-2001.  
XX  
PP 10-NOV-2000; 2000WO-US30792.  
XX  
PR 12-NOV-1999; 99US-0439058.  
PR 10-NOV-2000; 2000US-0439058.

XX  
PA (FIBR-) FIBROGEN INC.  
XX  
PT Bell MP, Neff TB, Polarek JW, Seeley TW;  
XX  
WPI: 2001-335911/35.  
DR N-PSDB; AAD06578.

XX  
PT Novel isolated and purified bovine or porcine collagens and gelatins  
PT useful in medical, pharmaceutical, food and cosmetic industries, as  
PT vaccine, and for treating autoimmune disorders, infections and cancer  
XX  
PS Example 5; Fig 12; 168pp; English.

XX  
CC The present sequence is porcine alpha(III) collagen. The present  
CC invention relates to recombinant synthesis of collagens and gelatins  
CC derived from animals. Collagen is useful in medical, pharmaceutical,  
CC food and cosmetic industries. Collagen is an important component of  
CC arterial sealants, bone grafts, drug delivery system, dermal implants,  
CC haemostats, and incontinence implants, and for treating autoimmune  
CC disorders such as rheumatoid arthritis. Collagen is useful in food  
CC products such as sausage casings, and in cosmetics or facial and skin  
CC formulations for treating viral infections, autoimmune diseases and  
CC cancer. Gelatin is useful in the manufacture or as a component of  
CC various pharmaceutical and medical devices and products, in food and  
CC beverage industries, in hair care and skin care products, as a glue or  
CC adhesive in various manufacturing processes, as a light-sensitive coating  
CC processes, in printing and photographic applications, in laboratory  
CC application, and as a component in various gels used for biochemical and  
CC electrophoretic analysis, including enzymographic gels.  
XX

SO Sequence 1466 AA;

Query Match 58.7%; Score 54; DB 22; Length 146;  
Best Local Similarity 66.7%; Pred. No. 33; Mismatches 0; Indels 5; Gaps 0;  
Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 PAGPWGPNGKCKVG 15  
Db 1142 PVGPGSPPGKGASG 1156

RESULT 29  
ABG15191

ID ABG15191 standard; Protein: 1469 AA.

XX AC ABG15191;

XX DT 18-FEB-2002 (first entry)

XX DE Novel human diagnostic protein #15182.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
OS Homo sapiens.

PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PP 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

DR WPI: 2001-639362/73.  
DR N-PSDB; AAST9378.

XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -

XX  
PS Claim 20; SEQ ID NO 45550; 103pp; English.

XX  
CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC and gene mapping, and in recombinant production of (II). The  
CC poly nucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABC01001-ABG0377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

SO Sequence 1469 AA;

Query Match 58.7%; Score 54; DB 22; Length 1469;  
Best Local Similarity 66.7%; Pred. No. 33; Mismatches 0; Indels 5; Gaps 0;  
Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 PAGPWGPNGKCKVG 15  
Db 1144 PVGPGSPPGKGCTSG 1158

RESULT 30

AAR53257  
ID AAR53257 standard; Protein: 1838 AA.

XX AC AAR53257;

XX DT 12-JAN-1995 (first entry)

XX DE Human collagen (Type V).

XX KW Human collagen; alpha 1; v type collagen; Placental mRNA.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers  
FT Peptide 1..38  
FT /label= signal-peptide

FT Region 44..538  
FT /note= "contains (Gly-X-Y) repeats"

FT Binding-site 645..647  
FT /label= RGD

FT /note= "cell adhesion motif"

FT Binding-site 663..665  
FT /label= RGD

FT /note= "cell adhesion motif"

FT Domain 897..933  
FT /label= heparin\_binding\_domain

FT Region 1573..1838  
FT /label= C-terminal\_region  
FT /note= "contains 8 Cys residues"

XX  
PN JP06105687-A.  
XX  
PD 19-APR-1994.  
XX  
PP 27-DEC-1991; 91JP-0358300.  
PR 27-DEC-1991; 91JP-0358300.  
XX  
PA (TAKI ) TAKARA SHUZO CO LTD.  
XX  
DR WPI; 1994-163129/20.  
DR N-PSDB; AAQ64556.  
XX  
PT Human collagen V-type gene - is used for diagnosis of human  
PT collagen V-type related diseases  
XX  
PS Claim 1; Page 6-14; 19PP; Japanese.  
XX  
CC This amino acid sequence of type V collagen contains several distinct  
CC domains including a region comprising repeated (Gly-X-Y) motifs and a  
CC central domain containing two copies of the Arg-Gly-Asp cell adhesion  
motif. The cDNA sequence encoding type V collagen was isolated from a  
CC human placental library and will be useful for diagnosis of diseases  
CC related to type V collagen.  
XX  
SQ Sequence 1838 AA;  
Query Match 58.7%; Score 54; DB 15; Length 1838;  
Best Local Similarity 66.7%; Pred. No. 41;  
Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
QY 1 PAGWGPNKGDKWIG 15  
Db 1145 PAGWGPNKGDKWIG 1159

Search completed: November 1, 2002, 12:52:35  
Job time : 28.5 secs

Copyright (c) 1993 - 2002 Compugen Ltd.

GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: November 1, 2002, 12:52:42 ; Search time 10.5 seconds

(without alignments) 34.894 Million cell updates/sec

Title: US-09-529-691a-3

Perfect score: 92

Sequence: 1 PAGPGNGPKDGKV 15

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Issued Patents AA:\*

1: /cgn2\_5/podata/1/1aa/5A\_COMB.pep:\*

2: /cgn2\_6/podata/1/1aa/5B\_COMB.pep:\*

3: /cgn2\_6/podata/1/1aa/5A\_COMB.pep:\*

4: /cgn2\_6/podata/1/1aa/5B\_COMB.pep:\*

5: /cgn2\_6/podata/1/1aa/POTUS\_COMB.pep:\*

6: /cgn2\_6/podata/1/1aa/backfiles1.pep:\*

Pred. No. 1 is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

**SUMMARIES**

Result No.	Score	Query Match	Length	DB ID	Description
1	63	60.5	534	4	US-09-029-348-5
2	63	68.5	535	4	US-09-029-348-1
3	63	68.5	537	4	US-09-029-348-4
4	63	68.5	1024	3	US-08-931-820-2
5	63	68.5	1366	4	US-09-570-573-19
6	63	68.5	1366	4	US-09-548-608-19
7	63	68.5	1418	4	US-09-010-999-1
8	59	64.1	1442	2	US-08-316-650-12
9	57	62.0	1442	5	PCT-US95-02251-12
10	57	62.0	41	1	US-08-395-816-2
11	55	59.8	54	2	US-08-316-650-14
12	55	59.8	54	5	PCT-US95-02251-14
13	55	59.8	1060	3	US-08-931-820-3
14	55	59.8	1418	3	US-08-963-825-20
15	55	59.8	1418	4	US-09-573-20
16	55	59.8	1418	4	US-09-548-608-20
17	55	59.8	1057	3	US-08-931-820-4
18	54	58.7	1078	3	US-08-963-825-21
19	54	58.7	1078	4	US-09-573-21
20	54	58.7	1078	4	US-09-548-608-21
21	54	58.7	1057	3	US-08-911-820-1
22	52	56.5	1341	3	US-08-963-825-18
23	52	56.5	1341	4	US-09-570-573-18
24	52	56.5	1341	4	US-09-608-18
25	52	56.5	1341	4	US-09-951-565-1
26	51	55.4	26	1	US-08-246-242-9
27	55.4		26	5	PCT-US96-00206-1

**ALIGNMENTS**

RESULT 1

US-09-029-348-5

; Sequence 5, Application US/09029348

; Patent No. 6171827

; GENERAL INFORMATION:

; APPLICANT: THE VICTORIA UNIVERSITY OF MANCHESTER

; TITLE OF INVENTION: NOVEL PROCOLLAGENS

; FILE REFERENCE: d087857pus LISTING

; CURRENT APPLICATION NUMBER: US/09/029, 348

; CURRENT FILING DATE: 1998-05-07

; NUMBER OF SEQ ID NOS: 20

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 5

; LENGTH: 534

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: SEQUENCE

; OTHER INFORMATION: DERIVED FROM cDNA OF PROCOLLAGENS

US-09-029-348-5

Query Match 68.5%; Score 63; DB 4; Length 534;

Best Local Similarity 73.3%; Pred. NO. 0.18; 3; Indels 0; Gaps 0;

Matches 11; Conservative 1; Mismatches 3;

Qy 1 PAGPGNGPKDGKV 15

Db 223 PAGPGNGPKDGKV 237

RESULT 2

US-09-029-348-1

; Sequence 1, Application US/09029348

; Patent No. 6171827

; GENERAL INFORMATION:

; APPLICANT: THE VICTORIA UNIVERSITY OF MANCHESTER

; TITLE OF INVENTION: NOVEL PROCOLLAGENS

; FILE REFERENCE: d087857pus LISTING

; CURRENT APPLICATION NUMBER: US/09/029, 348

; CURRENT FILING DATE: 1998-05-07

; NUMBER OF SEQ ID NOS: 20

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 1

; LENGTH: 535

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: SEQUENCE

; OTHER INFORMATION: DERIVED FROM cDNA OF PROCOLLAGENS

; US-09-029-348-1

Query Match 68.5%; Score 63; DB 4; Length 535;  
Best Local Similarity 73.3%; Pred. No. 0.18; Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 PAGPWNPGNGDKGVG 15  
Db 224 PAGPSGPAGKGRTG 238

---

RESULT 3 US-09-029-348-4

Sequence 4, Application US/09029348  
Patient No. 6171827

GENERAL INFORMATION:

APPLICANT: THE VICTORIA UNIVERSITY OF MANCHESTER  
TITLE OF INVENTION: NOVEL PROCOLLAGENS  
FILE REFERENCE: Q087857PUS LISTING

CURRENT APPLICATION NUMBER: US/09/029, 348  
CURRENT FILING DATE: 1998-05-07  
NUMBER OF SEQ ID NOS: 20

SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 4  
LENGTH: 537  
TYPE: PRT  
ORGANISM: Artificial sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: SEQUENCE

US-09-029-348-4

Query Match 68.5%; Score 63; DB 4; Length 537;  
Best Local Similarity 73.3%; Pred. No. 0.18; Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 PAGPWNPGNGDKGVG 15  
Db 224 PAGPSGPAGKGRTG 238

---

RESULT 4 US-08-931-820-2

Sequence 2, Application US/08931820  
Patient No. 6010863

GENERAL INFORMATION:

APPLICANT:

TITLE OF INVENTION: Assay for collagen degradation  
NUMBER OF SEQUENCES: 4

COMPUTER READABLE FORM:

COMPUTER TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/963, 825  
FILING DATE: 21-JAN-1994  
CLASSIFICATION: 436  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US/08/187, 319  
FILING DATE: 21-JAN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Gogoris, Adda C.  
REGISTRATION NUMBER: 29, 714  
REFERENCE/DOCKET NUMBER: 4305/08701  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-527-7700  
TELEFAX: 212-753-6237  
TELEX: 236637

INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1365 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
ORIGINAL SOURCE:  
IMMEDIATE SOURCE:  
CLONE: collagen alpha 2- type I

US-08-963-825-19

Query Match 68.5%; Score 63; DB 3; Length 1366;  
Best Local Similarity 73.3%; Pred. No. 0.49; Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 PAGPWNPGNGDKGVG 15  
Db 1055 PAGPSGPAGKGRTG 1069

---

RESULT 5 US-08-963-825-19

Sequence 19, Application US/08963825  
Patient No. 6110689

GENERAL INFORMATION:

APPLICANT: Oqvist, Per  
TITLE OF INVENTION: A Method for Assaying Collagen Fragments in Body Fluids, A Test Kit and Means for Carrying Out the Method and Use of the Method to Diagnose the Presence of Disorders Associated with the Metabolism of Collagen

TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of Disorders Associated with the Metabolism of Collagen

NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Darby & Darby PC  
STREET: 805 Third Avenue  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10022  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/963, 825  
FILING DATE:  
CLASSIFICATION: 436  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US/08/187, 319  
FILING DATE: 21-JAN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Gogoris, Adda C.  
REGISTRATION NUMBER: 29, 714  
REFERENCE/DOCKET NUMBER: 4305/08701  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-527-7700  
TELEFAX: 212-753-6237  
TELEX: 236637

INFORMATION FOR SEQ ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1365 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
ORIGINAL SOURCE:  
IMMEDIATE SOURCE:  
CLONE: collagen alpha 2- type I

US-08-963-825-19

Query Match 68.5%; Score 63; DB 3; Length 1366;  
Best Local Similarity 73.3%; Pred. No. 0.49; Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 PAGPWNPGNGDKGVG 15  
Db 1055 PAGPSGPAGKGRTG 1069

---

RESULT 6 US-09-570-573-19

Sequence 19, Application US/09570573  
Patient No. 6342361

US-08-931-820-2

## GENERAL INFORMATION:

APPLICANT: Qvist, Per

APPLICANT: Bonde, Martin

TITLE OF INVENTION: A Method for Assaying Collagen Fragments

TITLE OF INVENTION: In Body Fluids, A Test Kit and Means for Carrying Out the

TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of

TITLE OF INVENTION: Disorders Associated with the Metabolism of

NUMBER OF SEQUENCES: 21

CORRESPONDENCE ADDRESS:

ADDRESSEE: Darby &amp; Darby PC

STREET: 805 Third Avenue

CITY: New York

STATE: New York

COUNTRY: USA

ZIP: 10022

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/548,608

FILING DATE:

CLASSIFICATION:

PRIORITY APPLICATION DATA:

APPLICATION NUMBER: 08/187,319

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Goboris, Adda C

REGISTRATION NUMBER: 29,714

REFERENCE/DOCKET NUMBER: 4305/08701

TELECOMMUNICATION INFORMATION:

TELEPHONE: 212-527-7700

TELEFAX: 212-527-6237

TELEX: 216617

INFORMATION FOR SEQ ID NO: 19:

SEQUENCE CHARACTERISTICS:

LENGTH: 1366 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: Protein

ORIGINAL SOURCE:

ORGANISM: Homo sapiens

IMMEDIATE SOURCE:

CLONE: collagen alpha 2- type I

US-09-548-608-19

Query Match Best Local Similarity 68.5%; Score 63; DB 4; Length 1366;

Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 PAGPWGNGKDKVKG 15

Db 1055 PAGPSGPAGDKRTG 1069

RESULT 8

US-09-010-999-1

Sequence 1, Application US/09010999

PATENT NO. 6132076

GENERAL INFORMATION:

APPLICANT: Pole, Anthony R.

APPLICANT: Hollander, Anthony P.

APPLICANT: Billinghamhurst, R. C.

TITLE OF INVENTION: IMMUNOASSAYS FOR THE MEASUREMENT OF

TITLE OF INVENTION: COLLAGEN DENATURATION AND CLEAVAGE IN CARTILAGE

NUMBER OF SEQUENCES: 16

CORRESPONDENCE ADDRESS:

ADDRESSEE: Foley &amp; Lardner

STREET: 3000 K Street, N.W., Suite 500

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20007-5109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/010,999

FILING DATE: 22-JAN-1998

CLASSIFICATION: 435

PRIORITY APPLICATION DATA:

APPLICATION NUMBER: US 08/448,501

FILING DATE: 17-JUL-1995

RESULT 7

US-09-448-608-19

Sequence 19, Application US/09548608

PATENT NO. 635442

GENERAL INFORMATION:

APPLICANT: Qvist, Per

APPLICANT: Bonde, Martin

TITLE OF INVENTION: A Method for Assaying Collagen Fragments

TITLE OF INVENTION: In Body Fluids, A Test Kit and Means for Carrying Out the

TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of

TITLE OF INVENTION: Disorders Associated with the Metabolism of

NUMBER OF SEQUENCES: 21

CORRESPONDENCE ADDRESS:

ADDRESSEE: Darby &amp; Darby PC

STREET: 805 Third Avenue

CITY: New York

STATE: New York

PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/984,123  
 FILING DATE: 04-DEC-1992  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Bent, Stephen A.  
 REGISTRATION NUMBER: 29,768  
 REFERENCE/DOCKET NUMBER: 032931/0212  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (202)672-5300  
 TELEX: 904136  
 INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1418 amino acids  
 TYPE: amino acid  
 STRANDEDNESS:  
 TOPOLOGY: linear  
 MOLECULE TYPE: peptide  
 ORIGINAL SOURCE:  
 ORGANISM: Human Type II collagen  
 US-09-010-999-1

Query Match 64.1%; Score 59; DB 4; Length 1418;  
 Best Local Similarity 66.7%; Pred. No. 1.8; Matches 10;  
 保守型 2; Mismatches 3; Indels 0; Gaps 0;  
 Qy 1 PAGPWGPNGKDGKVG 15  
 Db 1096 PPGPGVGPCKDGAMG 1110

RESULT 9  
 US-08-316-650-12  
 Sequence 12, Application US/08316650  
 ;  
 Patent No. 594296  
 GENERAL INFORMATION:  
 APPLICANT: Bonadio, Jeffrey  
 APPLICANT: Roessler, Blake J.  
 APPLICANT: Goldstein, Steven A.  
 APPLICANT: Lin, Wushan  
 TITLE OF INVENTION: METHODS AND COMPOSITIONS  
 TITLE OF INVENTION: FOR STIMULATING BONE CELLS  
 NUMBER OF SEQUENCES: 15  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Arnold, White & Durkee  
 STREET: P.O. Box 4433  
 CITY: Houston  
 STATE: Texas  
 COUNTRY: United States of America  
 ZIP: 77210  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII  
 SOFTWARE: Patent Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: PCT/US95/02251  
 FILING DATE: CONCURRENTLY HEREWITH  
 CLASSIFICATION:  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/316,650  
 FILING DATE: 30-SEP-1994  
 CLASSIFICATION:  
 APPLICATION NUMBER: US 08/199,780  
 FILING DATE: 18-FEB-1994  
 CLASSIFICATION:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Parker, David L.  
 REGISTRATION NUMBER: 32,165  
 REFERENCE/DOCKET NUMBER: UMIC009P--  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (512) 418-3000  
 TELEX: (713) 789-2679  
 FILING DATE: 30-SEP-1994  
 CLASSIFICATION: 514  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/199,780  
 FILING DATE: 30-SEP-1994  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Parker, David L.  
 REGISTRATION NUMBER: 32,165  
 REFERENCE/DOCKET NUMBER: UMIC:008  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (512) 418-3000  
 TELEX: (713) 789-2679  
 INFORMATION FOR SEQ ID NO: 12:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1442 amino acids  
 TYPE: amino acid  
 STRANDEDNESS:  
 TOPOLOGY: linear  
 MOLECULE TYPE: peptide  
 ORIGINAL SOURCE:  
 ORGANISM: Human Type II collagen  
 US-09-010-999-1

RESULT 10  
 PCT-US95/02251-12  
 Sequence 12, Application PC/TUS9502251  
 ;  
 GENERAL INFORMATION:  
 APPLICANT: METHODS AND COMPOSITIONS FOR STIMULATING BONE  
 TITLE OF INVENTION: CELLS  
 NUMBER OF SEQUENCES: 18  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Arnold, White & Durkee  
 STREET: P.O. Box 4433  
 CITY: Houston  
 STATE: Texas  
 COUNTRY: United States of America  
 ZIP: 77210  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII  
 SOFTWARE: Patent Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: PCT/US95/02251  
 FILING DATE: CONCURRENTLY HEREWITH  
 CLASSIFICATION:  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/316,650  
 FILING DATE: 30-SEP-1994  
 CLASSIFICATION:  
 APPLICATION NUMBER: US 08/199,780  
 FILING DATE: 18-FEB-1994  
 CLASSIFICATION:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Parker, David L.  
 REGISTRATION NUMBER: 32,165  
 REFERENCE/DOCKET NUMBER: UMIC009P--  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (512) 418-3000  
 TELEX: (713) 789-2679  
 FILING DATE: 30-SEP-1994  
 CLASSIFICATION: 514  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/199,780  
 FILING DATE: 30-SEP-1994  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Parker, David L.  
 REGISTRATION NUMBER: 32,165  
 REFERENCE/DOCKET NUMBER: UMIC:008  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (512) 418-3000  
 TELEX: 79-0924  
 INFORMATION FOR SEQ ID NO: 12:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1442 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: peptide  
 ORIGINAL SOURCE:  
 ORGANISM: Human Type II collagen  
 US-08-316-650-12

Query Match 62.0%; Score 57; DB 5; Length 1442;  
 Best Local Similarity 73.3%; Pred. No. 3.6; Matches 11; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 Qy 1 PAGPWGPNGKDGKVG 15  
 Db 871 PAGPPGPAGKDGPKG 885

RESULT 11  
US 08-395-816-2  
Sequence 2, Application US/08395816  
; Patent No. 574751  
GENERAL INFORMATION:  
APPLICANT: Toshiki TANAKA et al.  
TITLE OF INVENTION: PEPTIDE DERIVATIVES HAVING  
TITLE OF INVENTION: BINDING ACTIVITY TO MODIFIED LOW DENSITY LIPOPROTEIN  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Wenderoth, Lind & Ponack  
STREET: 805 Fifteenth Street, N.W., #700  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 5 1/4 inch, 500 kb storage  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: Wordperfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/395,816  
APPLICATION DATE: FILING DATE: February 28, 1995  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Warren M. Cheek, Jr.  
REGISTRATION NUMBER: 33,367  
REFERENCE/DOCKET NUMBER:  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-371-8850  
TELEFAX:  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 41 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLogy: linear  
MOLECULE TYPE: peptide  
FEATURE:  
NAME/KEY: modified site  
LOCATION: 4, 7, 10, 13, 34, 37, 40  
IDENTIFICATION METHOD:  
OTHER INFORMATION: /note:- "4 Hyp"

Query Match 59.8%; Score 55; DB 1; Length 41;  
Best Local Similarity 66.7%; Pred. No. 0.16; 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 PAGGWPNGKDGKVG 15  
| || || || || |  
Db 6 PXGXPXPKTGKG 20

---

RESULT 12  
US 08-316-650-14  
Sequence 14, Application US/08316650  
; Patent No. 5942496  
GENERAL INFORMATION:  
APPLICANT: Jeffrey Bonadio  
APPLICANT: Roessler, Blake J.  
APPLICANT: Goldstein, Steven A.  
APPLICANT: Lin, Wushan  
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR STIMULATING BONE  
TITLE OF INVENTION: FOR STIMULATING BONE CELLS  
NUMBER OF SEQUENCES: 15  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Arnold, White & Durkee  
STREET: P.O. Box 4433  
CITY: Houston  
STATE: Texas  
ZIP: 77210  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/02251  
FILING DATE: CONCURRENTLY HEREWITH  
CLASSIFICATION:  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 08/316,650  
FILING DATE: 30-SEP-1994  
CLASSIFICATION:  
APPLICATION NUMBER: US 08/199,780  
FILING DATE: 18-FEB-1994  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:

COUNTRY: USA  
ZIP: 77210  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/316,650  
FILING DATE: 30-SEP-1994  
CLASSIFICATION:  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 08/199,780  
FILING DATE: 30-SEP-1994  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:

COUNTRY: USA  
ZIP: 77210  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/316,650  
FILING DATE: 30-SEP-1994  
CLASSIFICATION:  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 08/199,780  
FILING DATE: 18-FEB-1994  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:

COUNTRY: USA  
ZIP: 77210  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/316,650  
FILING DATE: 30-SEP-1994  
CLASSIFICATION:  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 08/199,780  
FILING DATE: 30-SEP-1994  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:

NAME: Parker, David L.  
 REGISTRATION NUMBER: 32,165  
 REFERENCE/DOCKET NUMBER: UNIC009P--  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (512) 418-3000  
 TELEX: 79-0224  
 INFORMATION FOR SEQ ID NO: 14:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 54 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: peptide  
 PCT-US95-02251-14

Query Match 59.8%; Score 55; DB 5; Length 54;  
 Best Local Similarity 66.7%; Pred. No. 0.22; 4; Indels 0; Gaps 0;  
 Matches 10; Conservative 1; Mismatches 4;

Qy 1 PAGPMPNGKDGKVG 15  
 Db 20 PPGPVGPGSKDGANG 34

RESULT 14  
 US-08-931-820-3  
 Sequence 3, Application US/08931820  
 Patent No. 6010863  
 GENERAL INFORMATION:

APPLICANT:  
 TITLE OF INVENTION: Assay for collagen degradation  
 NUMBER OF SEQUENCES: 4  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/931,820  
 FILING DATE:  
 CLASSIFICATION: 436  
 PRIORITY APPLICATION DATA:  
 APPLICATION NUMBER: US/08/187,319  
 FILING DATE: 21-JAN-1994  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Gogoris, Adda C  
 REGISTRATION NUMBER: 29,714  
 REFERENCE/DOCKET NUMBER: 4305/08701  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 212-527-7700  
 TELEX: 236687

INFORMATION FOR SEQ ID NO: 20:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1418 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 ORIGINAL SOURCE:  
 ORGANISM: Homo sapiens  
 IMMEDIATE SOURCE:  
 CLONE: COLLAGEN - ALPHA 1 (III)  
 US-08-963-825-20

Query Match 59.8%; Score 55; DB 3; Length 1418;  
 Best Local Similarity 66.7%; Pred. No. 6.7; 4; Indels 0; Gaps 0;  
 Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 1 PAGPMPNGKDGKVG 15  
 Db 1096 PPGPVGPGSKDGANG 1110

RESULT 16  
 US-09-570-573-20  
 Sequence 20, Application US/09570573  
 Patent No. 6342361  
 GENERAL INFORMATION:

APPLICANT: Qvist, Per  
 APPLICANT: Bond, Martin  
 TITLE OF INVENTION: A Method for Assaying Collagen Fragments Carrying Out the  
 TITLE OF INVENTION: In Body Fluids, A Test Kit and Means for Carrying Out the  
 TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence or  
 TITLE OF INVENTION: Disorders Associated with the Metabolism of  
 NUMBER OF SEQUENCES: 21

CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Darby & Darby PC  
 STREET: 805 Third Avenue  
 CITY: New York  
 STATE: New York  
 COUNTRY: USA  
 ZIP: 10022

APPLICANT: Bond, Martin  
 TITLE OF INVENTION: A Method for Assaying Collagen Fragments Carrying Out the  
 TITLE OF INVENTION: In Body Fluids, A Test Kit and Means for Carrying Out the  
 TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence or  
 TITLE OF INVENTION: Disorders Associated with the Metabolism of  
 NUMBER OF SEQUENCES: 21

CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Darby & Darby PC  
 STREET: 805 Third Avenue  
 CITY: New York  
 STATE: New York  
 COUNTRY: USA  
 ZIP: 10022

APPLICANT: Bond, Martin  
 TITLE OF INVENTION: A Method for Assaying Collagen Fragments Carrying Out the  
 TITLE OF INVENTION: In Body Fluids, A Test Kit and Means for Carrying Out the  
 TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence or  
 TITLE OF INVENTION: Disorders Associated with the Metabolism of  
 NUMBER OF SEQUENCES: 21

CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Darby & Darby PC  
 STREET: 805 Third Avenue  
 CITY: New York  
 STATE: New York  
 COUNTRY: USA  
 ZIP: 10022

RESULT 15  
 US-08-963-825-20  
 Sequence 20, Application US/08933825  
 Patent No. 6110689  
 GENERAL INFORMATION:  
 APPLICANT: Qvist, Per

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/548,608  
 FILING DATE:  
 CLASSIFICATION:  
 PRIORITY APPLICATION DATA:  
 APPLICATION NUMBER: 08/187,319  
 FILING DATE:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Gogoris, Adda C  
 REGISTRATION NUMBER: 29,714  
 REFERENCE/DOCKET NUMBER: 4305/08701  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 212-527-7700  
 TELEFAX: 212-753-6237  
 TELEX: 236887  
 INFORMATION FOR SEQ ID NO: 20:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1418 amino acids  
 LENGTH: 1418 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 ORIGINAL SOURCE:  
 ORGANISM: Homo sapiens  
 IMMEDIATE SOURCE:  
 CLONE: COLLAGEN -ALPHA 1 (II)  
 CLONE: COLLAGEN -ALPHA 1 (II)  
 RESULT 17  
 US-09-548-608-20  
 Sequence 20, Application US/09548608  
 Patient No. 6355412  
 GENERAL INFORMATION:  
 APPLICANT: Qrist, Per  
 APPLICANT: Bonde, Martin  
 TITLE OF INVENTION: A Method for Assaying Collagen Fragments  
 TITLE OF INVENTION: In Body Fluids, A Test Kit and Means for Carrying Out the  
 TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of  
 TITLE OF INVENTION: Disorders Associated with the Metabolism of  
 NUMBER OF SEQUENCES: 21  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Darby & Darby PC  
 STREET: 805 Third Avenue  
 CITY: New York  
 STATE: New York  
 COUNTRY: USA  
 ZIP: 10022  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/548,608  
 FILING DATE:  
 CLASSIFICATION: 435  
 PRIORITY APPLICATION DATA:  
 APPLICATION NUMBER: EP 96202596.1  
 FILING DATE:  
 INFORMATION FOR SEQ ID NO: 4:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1057 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 HYPOTHETICAL: NO  
 ORIGINAL SOURCE:  
 ORGANISM: Homo sapiens  
 TISSUE TYPE: Collagen type III  
 FEATURE:  
 NAME/KEY: Modified-site  
 LOCATION: 1055  
 OTHER INFORMATION: /label- "Ala may be Pro"  
 ; OTHER INFORMATION: /note- "Ala may be Pro"  
 US-08-931-820-4  
 Query Match 58.7%; Score 54; DB 3; Length 1057;  
 Best Local Similarity 66.7%; Pred No. 6.8;  
 Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 Qy 1 PAGCAGPCKDGKVG 15  
 Db 1096 PPGPVGPGSKDGANG 1110  
 RESULT 18  
 US-08-931-820-4  
 Sequence 4, Application US/08931820  
 Patient No. 6010863  
 GENERAL INFORMATION:  
 APPLICANT:  
 TITLE OF INVENTION: Assay for collagen degradation  
 NUMBER OF SEQUENCES: 4  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, version #1.25 (EPO)  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/931,820  
 FILING DATE:  
 CLASSIFICATION: 435  
 PRIORITY APPLICATION DATA:  
 APPLICATION NUMBER: EP 96202596.1  
 FILING DATE:  
 INFORMATION FOR SEQ ID NO: 4:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1057 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 HYPOTHETICAL: NO  
 ORIGINAL SOURCE:  
 ORGANISM: Homo sapiens  
 TISSUE TYPE: Collagen type III  
 FEATURE:  
 NAME/KEY: Modified-site  
 LOCATION: 1055  
 OTHER INFORMATION: /label- "Ala may be Pro"  
 ; OTHER INFORMATION: /note- "Ala may be Pro"  
 US-08-931-820-4  
 Query Match 58.7%; Score 54; DB 3; Length 1057;  
 Best Local Similarity 66.7%; Pred No. 6.8;  
 Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 Qy 1 PAGCAGPCKDGKVG 15  
 Db 993 PVGPSPGPKDGTSQ 1007

RESULT 19  
US-08-963-825-21  
Sequence 21, Application US/08963825  
Patent No. 6110689  
GENERAL INFORMATION:  
APPLICANT: Oqvist, Per  
APPLICANT: Bonde, Martin  
TITLE OF INVENTION: A Method for Assaying Collagen Fragments  
TITLE OF INVENTION: In Body Fluids, A Test Kit and Means for Carrying Out the  
TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of  
TITLE OF INVENTION: Disorders Associated with the Metabolism of  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Darby & Darby PC  
STREET: 805 Third Avenue  
CITY: New York  
STATE: New York  
COUNTRY: USA  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10022  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/570,573  
FILING DATE:  
CLASSIFICATION:  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US/08/187,319  
FILING DATE: 21-JAN-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Gogoris, Adda C  
REGISTRATION NUMBER: 29,714  
REGISTRATION NUMBER: 4305/08701  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-527-7700  
TELEX: 236687  
INFORMATION FOR SEQ ID NO: 21:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1078 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
ORIGINAL SOURCE:  
ORGANISM: Homo sapiens  
IMMEDIATE SOURCE:  
CLONE: COLLAGEN ALPHA 1 (III)  
; US-08-963-825-21  
Query Match 58.7%; Score 54; DB 3; Length 1078;  
Best Local Similarity 66.7%; Pred. No. 6.9;  
Matches 10; Conservative 66.7%; Mismatches 5; Indels 0; Gaps 0;  
QY 1 PAGPWPPNGKDGKVG 15  
Db 994 PVGPSSGPPKGDTSG 1008  
; US-09-570-573-21  
Query Match 58.7%; Score 54; DB 4; Length 1078;  
Best Local Similarity 66.7%; Pred. No. 6.9;  
Matches 10; Conservative 66.7%; Mismatches 5; Indels 0; Gaps 0;  
QY 1 PAGPWPPNGKDGKVG 15  
Db 994 PVGPSSGPPKGDTSG 1008  
; RESULT 21  
US-09-548-608-21  
Sequence 21, Application US/09548608  
Patent No. 6355442  
GENERAL INFORMATION:  
APPLICANT: Oqvist, Per  
APPLICANT: Bonde, Martin  
TITLE OF INVENTION: A Method for Assaying Collagen Fragments  
TITLE OF INVENTION: In Body Fluids, A Test Kit and Means for Carrying Out the  
TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of  
TITLE OF INVENTION: Disorders Associated with the Metabolism of  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Darby & Darby PC  
STREET: 805 Third Avenue  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10022  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/548,608  
 FILING DATE: 1994-01-12  
 CLASSIFICATION: 436  
 PRIORITY APPLICATION DATA:  
 APPLICATION NUMBER: 08/187,319  
 FILING DATE:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Gogoris, Adda C  
 REGISTRATION NUMBER: 29,714  
 REFERENCE/DOCKET NUMBER: 4305/08701  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 212-527-7700  
 TELEX: 236687  
 INFORMATION FOR SEQ ID NO: 21:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1078 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 ORIGINAL SOURCE: Homo sapiens  
 IMMEDIATE SOURCE:  
 CLONE: COLLAGEN ALPHA 1 (III)  
 US-09-548-608-21

RESULT 23  
 Query Match 58.7%; Score 54; DB 4; Length 1078;  
 Best Local Similarity 66.7%; Pred. No. 6.9;  
 Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 QY 1 PAGPWGPNKGDKGVG 15  
 ||||| ||| : |  
 Db 994 PVGPGSGPKDGMSG 1008

RESULT 22  
 US-08-931-820-1  
 Sequence 1, Application US/08931820  
 Patent No. 6010863  
 GENERAL INFORMATION:  
 APPLICANT:  
 TITLE OF INVENTION: Assay for collagen degradation  
 NUMBER OF SEQUENCES: 4  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, version #1.25 (EPO)  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/963,825  
 FILING DATE:  
 CLASSIFICATION: 436  
 PRIORITY APPLICATION DATA:  
 APPLICATION NUMBER: US/08/187,319  
 FILING DATE: 21-JAN-1994  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Gogoris, Adda C  
 REGISTRATION NUMBER: 29,714  
 REFERENCE/DOCKET NUMBER: 4305/08701  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 212-527-7700  
 TELEX: 236687  
 INFORMATION FOR SEQ ID NO: 18:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1341 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 ORIGINAL SOURCE: Homo sapiens  
 ORGANISM: Homo sapiens  
 IMMEDIATE SOURCE:  
 CLONE: COLLAGEN ALPHA 1 (I)  
 US-08-963-825-18

RESULT 24  
 Query Match 56.5%; Score 52; DB 3; Length 1341;  
 Best Local Similarity 60.0%; Pred. No. 17;  
 Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;  
 QY 1 PAGPWGPNKGDKGVG 15  
 ||||| ||| : |  
 Db 962 PAGPQGRGDKGETG 976

RESULT 24  
 US-09-570-573-18  
 Sequence 18, Application US/09570573  
 Patent No. 632361  
 GENERAL INFORMATION:  
 APPLICANT: Qvist, Per  
 APPLICANT: Bonde, Martin

Query Match 56.5%; Score 52; DB 3; Length 1057;  
 Best Local Similarity 60.0%; Pred. No. 13;  
 Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

TITLE OF INVENTION: A Method for Assaying Collagen Fragments  
 TITLE OF INVENTION: In Body Fluids, A Test Kit and Means for Carrying Out the  
 Method and Use of the Method to Diagnose the Presence of  
 TITLE OF INVENTION: Disorders Associated with the Metabolism of  
 NUMBER OF SEQUENCES: 21  
 CORRESPONDENCE ADDRESS:  
 ADDRESSE: Darby & Darby PC  
 STREET: 805 Third Avenue  
 CITY: New York  
 STATE: New York  
 COUNTRY: USA  
 ZIP: 10022

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/548,608

APPLICATION NUMBER: US/09/570,573  
 FILING DATE: 2/12/99  
 CLASSIFICATION:  
 PRIORITY APPLICATION DATA:  
 APPLICATION NUMBER: 08/187,319  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/570,573  
 FILING DATE: 2/12/99  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Gogoris, Adda C.  
 REGISTRATION NUMBER: 29,714  
 REFERENCE/DOCKET NUMBER: 4305/08701  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 212-527-7700  
 TELEFAX: 212-753-6237  
 TELEX: 236687

INFORMATION FOR SEQ ID NO: 18:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1341 amino acids  
 TYPE: amino acid

MOLECULE TYPE: protein  
 ORIGINAL SOURCE: Homo sapiens  
 ORGANISM: Homo sapiens  
 IMMEDIATE SOURCE:  
 CLONE: COLLAGEN ALPHA 1 (I)

US-09-570-573-18

TELEX: 236687

Query Match 56.5%; Score 52; DB 4; Length 1341;  
 Best Local Similarity 60.0%; Pred. No. 17; Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 PAGPWNPGNGDKGVG 15  
 1||| 1|| |: |  
 Db 962 PAGPQGPGRGDKGETG 976

RESULT 25

US-09-548-608-18

Sequence 1, Application US/07951565  
 Patent No. 5399347

GENERAL INFORMATION:  
 APPLICANT: Trentham, David E.

APPLICANT: Weinan, Howard L.

TITLE OF INVENTION: Method of Treating Rheumatoid Arthritis with  
 NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:  
 ADDRESSE: Darby & Darby  
 STREET: 805 Third Ave.  
 CITY: New York  
 STATE: New York  
 COUNTRY: U.S.A.  
 ZIP: 10022-7513

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/07/951,565  
 FILING DATE: 19920925  
 CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:  
 NAME: Gogoris, Adda C.  
 REGISTRATION NUMBER: 29,714  
 REFERENCE/DOCKET NUMBER: 1010/07300  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (212)527-7700  
 TELEFAX: (212)753-6237

COMPUTER READABLE FORM:

TELEX: 236687  
INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:  
LENGTH: 26 amino acids

TYPE: AMINO ACID

TOPOLOGY: linear

MOLECULE TYPE: peptide

HYPOTHETICAL: NO

FRAGMENT TYPE: internal

ORIGINAL SOURCE: Galius domesticus

ORGANISM: Galius domesticus

TISSUE TYPE: collagen

US-07-951-565-1

Query Match 55.4%; Score 51; DB 1; Length 26;  
Best Local Similarity 53.3%; Pred. No. 0.37; Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 PAGPWGNGKDGKV 15  
Db 1 PTGPLGPKGQIGELG 15

RESULT 27  
US-07-951-565-6

Sequence 6, Application US/07951565  
Patent No. 5309347

GENERAL INFORMATION:

APPLICANT: Trentham, David E.

ATTORNEY: Weiner, Howard L.

TITLE OF INVENTION: Method of Treating Rheumatoid Arthritis with

TITLE OF INVENTION: Type II Collagen

NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:

ADDRESSEE: Darby & Darby

STREET: 805 Third Ave.

CITY: New York

STATE: New York

ZIP: 10022-7513

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Goldstein, Jorge A.

REGISTRATION NUMBER: 29,021

REFERENCE/DOCKET NUMBER: 1383.0080000

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 371-2600

TELEFAX: (202) 371-2540

INFORMATION FOR SEQ ID NO: 9:

SEQUENCE CHARACTERISTICS:  
LENGTH: 26 amino acids

TYPE: amino acid

STRANDEDNESS: unknown

TOPOLOGY: unknown

MOLECULE TYPE: protein

US-08-246-242-9

Query Match 55.4%; Score 51; DB 1; Length 26;  
Best Local Similarity 53.3%; Pred. No. 0.37; Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 PAGPWGNGKDGKV 15  
Db 1 PTGPLGPKGQIGELG 15

RESULT 29  
PCN-US6-00206-1

Sequence 1, Application PC/TUS9600206

GENERAL INFORMATION:

APPLICANT: Immunologic Pharmaceutical Corporation

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR

TITLE OF INVENTION: TREATING RHEUMATOID ARTHRITIS

NUMBER OF SEQUENCES: 7

CORRESPONDENCE ADDRESS:

ADDRESSEE: Luppin & Kusner

STREET: 200 State Street

CITY: Boston

STATE: MA

COUNTRY: USA

ZIP: 02109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

Query Match 55.4%; Score 51; DB 1; Length 26;  
Best Local Similarity 53.3%; Pred. No. 0.37; Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 PAGPWGNGKDGKV 15

SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: PCT/US96/00206  
 FILING DATE:  
 CLASSIFICATION:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Kerner, Ann-Louise  
 REGISTRATION NUMBER: 33,523  
 REFERENCE/DOCKET NUMBER: IMZ-014PCT  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 617-466-6000  
 TELEFAX: 617-466-6040  
 INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 26 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: peptide  
 FRAGMENT TYPE: internal  
 ORIGINAL SOURCE: Bos taurus type II collagen  
 PCT-US96-00206-1

Query Match 55.4%; Score 51; DB 5; Length 26;  
 Best Local Similarity 53.3%; Pred. No. 0.37;  
 Matches 8; Conservative 3; Mismatches 4; Indels 0;  
 Gaps 0;

Qy 1 PAGPWGPNGDKGVG 15  
 Db 1 PTGPPIGPKGDKGEG 15

US-08-836-854-19 ; LENGTH: 464 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: peptide

US-08-836-854-19 ; Query Match 55.4%; Score 51; DB 2; Length 464;  
 Best Local Similarity 66.7%; Pred. No. 7.6; 5; Indels 0; Gaps 0;  
 Matches 10; Conservative 0; Mismatches 0;  
 Qy 1 PAGPWGPNGDKGVG 15  
 Db 442 PKGPPGPGDKGLPG 456

Search completed: November 1, 2002, 12:54:45  
 Job time : 11.5 secs

RESULT 30  
 US-08-836-854-19  
 Sequence 19 Application US/08836854  
 ;  
 Patent No. 5824547  
 GENERAL INFORMATION:  
 APPLICANT: HASHINO, Kinikazu  
 APPLICANT: MATSUSHITA, Hideyuki  
 APPLICANT: KATO, Ikunoshin  
 TITLE OF INVENTION: METHOD OF PRODUCTION OF TRANSFECTED CELLS  
 NUMBER OF SEQUENCES: 21  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Browdy and Neimark  
 STREET: 419 Seventh Street N.W. Ste. 300  
 CITY: Washington  
 STATE: D.C.  
 COUNTRY: USA  
 ZTP: 20004  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: FLOPPY disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/836,854  
 FILING DATE:  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: PCT/JP95/02425  
 FILING DATE: 29-NOV-1995  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 317721/1994  
 FILING DATE: 29-NOV-1994  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Browdy, Roger L.  
 REGISTRATION NUMBER: 25,618  
 REFERENCE/DOCKET NUMBER: HASHINO=1  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (202) 628-5197  
 TELEFAX: (202) 737-5528  
 INFORMATION FOR SEQ ID NO: 19:  
 SEQUENCE CHARACTERISTICS:

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

#### On protein - protein search, using sw model.

Run on: November 1, 2002, 12:38:32 ; Search time 13.5 Seconds  
(without alignments)  
106.766 Million cell updates/sec

Title: US-09-529-691a-3  
Perfect score: 92  
Sequence: 1 PAGPPGPNKGDKGVG 15

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

#### Searched:

283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 71:  
1: pir1:  
2: pir2:  
3: pir3:  
4: pir4:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	69	75.0	460	2 T33110	hypothetical protein C18H7.3 - Caenorhabditis elegans
2	66	71.7	291	2 T20942	hypothetical prote
3	63	68.5	1366	1 CGHU2S	hypothetical prote
4	60	65.2	1373	1 A42321	collagen alpha 2(I)
5	60	65.2	2944	2 A54849	collagen alpha 1(V)
6	57	62.0	365	2 S10847	collagen alpha 2(I)
7	57	62.0	1051	2 A35763	collagen alpha 2(C)
8	57	62.0	1419	2 A41182	collagen alpha 1(I)
9	57	62.0	1487	2 B41182	collagen alpha 1(I)
10	56	60.9	193	2 S07133	collagen alpha 1(I)
11	56	60.9	303	2 T19289	hypothetical prote
12	56	60.9	636	2 S41067	collagen alpha 1(I)
13	56	60.9	888	2 S28791	collagen alpha 1(X)
14	56	60.9	1075	2 T30842	serine-repeat anti
15	56	60.9	1486	1 B40333	collagen alpha 1(I)
16	56	60.9	1492	1 A40333	collagen alpha 1(I)
17	56	60.9	1805	1 CGHU2E	collagen alpha 1(X)
18	55	59.8	172	2 D41132	collagen-related p
19	55	59.8	326	2 T16841	hypothetical prote
20	55	59.8	363	2 T16831	hypothetical prote
21	55	59.8	464	2 S59513	collagen II AI pro
22	55	59.8	615	2 A52029	collagen alpha 1(I)
23	55	59.8	964	1 CGCH0S	collagen alpha 2(I)
24	55	59.8	1042	1 CGCH1S	collagen alpha 1(I)
25	55	59.8	1418	2 T45467	collagen alpha 1(I)
26	55	59.8	1464	2 S59856	collagen alpha 1(I)
27	55	59.8	1487	1 CGHU6C	collagen alpha 1(I)
28	55	59.8	1546	2 CGHU2E	collagen alpha 2(X)
29	58.7	228	1	A44982	collagen UC01 - p

#### ALIGNMENTS

##### RESULT 1

T33110 hypothetical protein C18H7.3 - Caenorhabditis elegans

C;Species: Caenorhabditis elegans  
C;Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 21-Jan-2000

R;Tin-Wollam, A.; Fronick, W.

submitted to the EMBL Data Library, May 1998

A; Description: The sequence of C. elegans cosmid C18H7.

A; Reference number: 221284

A; Accession: T33110

A; Status: preliminary; translated from GB/EMBL/DDJB

A; Molecule type: DNA

A; Residues: 1-460 <TIN>

A; Cross-references: EMBL:AF067607; PIDN: AAC17641.1; GSPDB:GN00022; CESP:C18H7.3

A; Experimental source: strain Bristol N2; clone C18H7

C; Genetics:

A; Gene: CESP:C18H7.3

A; Map position: 4

A; Introns: 84/1

C;Superfamily: Phaseolus glycine-rich cell wall protein 1.8

Query Match 75.0%; Score 69; DB 2; Length 460;  
Best Local Similarity 80.0%; Pred. No. 0.017; Mismatches 2; Indels 0; Gaps 0;

Matches 12; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 PAGPPGPNKGDKGV 15  
Db 237 PAGPPGPNKGDGEG 251

RESULT 2

T20942 hypothetical protein F15A2.1 - caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 18-Feb-2000

C;Accession: T20942

R;Gregory, J.

submitted to the EMBL Data Library, March 1996

A; Reference number: Z19349

A; Accession: T20942

A; Status: preliminary; translated from GB/EMBL/DDJB

A; Molecule type: DNA

A; Residues: 1-291 <WIL>

A; Cross-references: EMBL:Z70207; PIDN:CAA9128.1; GSPDB:GN00028; CESP:F15A2.1

A; Experimental source: clone F15A2

C; Genetics:

A; Gene: CESP:F15A2.1

A; Map position: X

A; Introns: 50/3

C;Superfamily: unassigned collagens

Query Match      71.7%;      Score 66;      DB 2;      Length 291;  
 Best Local Similarity      80.0%;      Pred. No. 0.029;      Mismatches 0;  
 Matches 12;      Conservative 0;      Mismatches 3;      Indels 0;      Gaps 0;

QY      1 PAGPWPNGPKDGKVG 15  
 Db      183 PAGPAGPEKGKDPVG 197

RESULT 3

CGH2S  
 collagen alpha 2(I) chain precursor - human  
 N;Alternate names: procollagen alpha 2(I) chain  
 C;Species: Homo sapiens (man)  
 C;Date: 30-Jun-1989 #sequence\_revision 25-Aug-1995 #text\_change 21-Jul-2000  
 C;Accession: A28500; S0824; S09176; I55311; A58111; A28472; R42165; A34405; A90567; I55  
 R;de Wet, W.; Bernard, M.; Benson-Chanda, V.; Chu, M.L.; Dickson, L.; Weil, D.; Ramirez,  
 J. Biol. Chem. 262, 16032-16036, 1987  
 A;Title: Organization of the human pro-alpha-2(I) collagen gene.  
 A;Reference number: A28500; MUID:88058962  
 A;Accession: A28500  
 A;Molecule type: DNA; mRNA  
 A;Residues: 1-248; N'; 250-1366 <DEN>  
 A;Cross-references: GB:J03464; NID:9170595; PIDN:AB53374.1; PID:9179596  
 R;Kuivaniemi, H.; Tromp, G.; Chu, M.L.; Prockop, D.J.  
 Biochem. J. 252, 633-640, 1988  
 A;Title: Structure of a full-length cDNA clone for the prepro-alpha-2(I) chain of human  
 A;Reference number: S00824; MUID:88339824  
 A;Accession: S00824  
 A;Molecule type: mRNA  
 A;Residues: 1-275; 'A', 277-332, 'V', 334-337, 'A', 339-482, 'A', 484-548, 'D', 550-765 <KU11>  
 A;Cross-references: EMBL:Y00724; NID:g30022; PIDN:CAA8709.1; PID:g30023  
 R;Dickson, L.A.; de Wet, W.; di Liberto, M.; Weil, D.; Ramirez, F.  
 Nucleic Acids Res. 13, 3427-3438, 1985  
 A;Title: Analysis of the promoter region and the N-propeptide domain of the human proalpha-1  
 A;Reference number: S09176; MUID:85242047  
 A;Accession: S09176  
 A;Molecule type: DNA  
 A;Residues: 1-23; 33-58, 'P', 60-93 <DIC>  
 A;Cross-references: EMBL:X02488; NID:g00998; PIDN:CAA26320.1; PID:g30099  
 R;Weil, D.; d'Alessio, M.; Ramirez, F.; Eyre, D.R.  
 J. Biol. Chem. 265, 16007-16011, 1990  
 A;Title: Structural and functional characterization of a splicing mutation in the pro-alpha-1  
 A;Accession: I55311; MUID:90368825  
 A;Status: translated from GB/EMBL/DDJB  
 A;Molecule type: DNA  
 A;Residues: 76-93 <WEI1>  
 A;Cross-references: GB:M35391; NID:9189684; PIDN:AAA60041.1; PID:g189685  
 A;Accession: A58111  
 A;Molecule type: protein  
 A;Residues: 23-75,94-96 <WEI2>  
 A;Note: mutant sequence from a patient with Ehlers-Danlos syndrome type VII  
 R;Wirtz, M.K.; Glanville, R.W.; Steinmann, B.; Rao, V.H.; Hollister, D.W.  
 J. Biol. Chem. 262, 16316-16385, 1987  
 A;Title: Ehlers-Danlos syndrome type VIIb. Deletion of 18 amino acids comprising the N-t  
 A;Reference number: A28472; MUID:88059013  
 A;Accession: A28472  
 A;Molecule type: protein  
 A;Residues: 32-75,94-111 <WIR>  
 A;Note: mutant sequence of patient with Ehlers-Danlos syndrome type VIIb  
 R;Chioto, A., Hockey, A.; Cole, W.G.  
 J. Biol. Chem. 267, 6361-6369, 1992  
 A;Title: A base substitution at the splice acceptor site of intron 5 of the COLIA2 gene  
 s-Danlos syndrome type VII.  
 A;Accession: A42165; MUID:92210617  
 A;Accession: A42165  
 A;Molecule type: mRNA  
 A;Residues: 50-126 <CH1>  
 A;Note: parts of this sequence were determined by protein sequencing; a mutant sequence  
 R;Weil, D.; d'Alessio, M.; Ramirez, F.; Steinmann, B.; Wirtz, M.K.; Glanville, R.W.; Hol  
 J. Biol. Chem. 264, 16804-16809, 1989

A;Title: Temperature-dependent expression of a collagen splicing defect in the fibroblast  
 A;Reference number: A34405; MUID:89380311  
 A;Accession: A34405  
 A;Status: not compared with conceptual translation  
 A;Molecule type: mRNA  
 A;Residues: 58-108 <WEI3>  
 A;Cross-references: GB:J05049  
 A;Note: the accession cited by the authors is not found in GenBank  
 A;Note: parts of this sequence were determined by protein sequencing; a mutant having  
 R;Click, B.M.; Bernstein, P.  
 Biochemistry 9, 4693-4705, 1970  
 A;Note: the compositions of peptides CNBR1, CNBR0, and CNBR2 were determined; evidence  
 R;Kuivaniemi, H.; Sabol, C.; Tromp, G.; Sippola-Thiele, M.; Prockop, D.J.  
 J. Biol. Chem. 263, 11407-11413, 1988  
 A;Title: Isolation and characterization of the cyanogen bromide peptides from the alp  
 A;Title: A 19-base pair deletion in the pro-alpha 2(I) gene of type I procollagen tha  
 A;Title: A 19-base pair deletion in the pro-alpha 2(I) gene of type I procollagen tha  
 A;Reference number: I55264; MUID:88298792  
 A;Accession: I55264  
 A;Status: translation not shown; translated from GB/EMBL/DDJB  
 A;Molecule type: DNA; mRNA  
 A;Residues: 145-197 <KU12>  
 A;Cross-references: GB:MA1671; NID:9189521; PIDN:AAA59994.1; PID:g5533606  
 A;Note: single base mutation in intron leads to abnormal splicing of mRNA  
 R;Chitman, S.D.; Shapiro, J.R.; McKinstry, M.B.; Stover, M.L.; Branson, P.; Rowe, D.W.  
 J. Bone Miner. Res. 7, 793-805, 1992  
 A;Title: Expression of mutant alpha 1(I)-procollagen in osteoclast and fibroblast cult  
 A;Reference number: I55485; MUID:92351816  
 A;Accession: I55485  
 A;Status: translated from GB/EMBL/DDJB  
 A;Molecule type: mRNA  
 A;Residues: 165-181; 200-213 <CH2>  
 A;Cross-references: GB:SA1099; NID:9253702; PIDN:AB22761.1; PID:g22703  
 A;Note: mutant sequence from a patient with osteogenesis imperfecta type IV  
 R;Morgan, P.H.; Jacobs, H.G.; Segrest, J.P.; Cunningham, L.W.  
 J. Biol. Chem. 245, 5042-5048, 1970  
 A;Title: Comparative study of glycopeptides derived from selected vertebrate collagen  
 A;Reference number: A92069; MUID:7100508  
 A;Accession: B92069  
 A;Molecule type: protein  
 A;Residues: 175-180 <MOR>  
 A;Experimental source: skin  
 A;Note: attachment of 2-O-alpha-D-glucosyl-D-beta-D-galactose to 5-hydroxylysine  
 R;Fietzek, P.P.; Furthmayr, H.; Kuehn, K.  
 Eur. J. Biochem. 47, 257-261, 1974  
 A;Title: Comparative sequence studies on alpha2-CB2 from calf, human, rabbit and pig  
 A;Reference number: A91224; MUID:75008198  
 A;Accession: A91224  
 A;Molecule type: protein  
 A;Residues: 418-447 <FIE>  
 R;Tromp, G.; Prockop, D.J.  
 Proc. Natl. Acad. Sci. U.S.A. 85, 5254-5258, 1988  
 A;Title: Single base mutation in the pro-alpha 2(I) collagen gene that causes efficie  
 A;Reference number: I59125; MUID:88276936  
 A;Accession: I59125  
 A;Status: translation not shown; translated from GB/EMBL/DDJB  
 A;Molecule type: DNA  
 A;Residues: 520-573 <TRO>  
 A;Cross-references: GB:M21352; NID:9180881; PIDN:AAA52053.1; PID:g180882  
 A;Note: single base mutation in exon 28  
 R;Bernard, M.P.; Myers, J.C.; Chu, M.L.; Ramirez, F.; Eikenberry, E.F.; Prockop, D.J.  
 Biochemistry 22, 1139-1145, 1983  
 A;Title: Structure of a cDNA for the proalpha-2 chain of human type I procollagen. Co  
 A;Reference number: S09174; MUID:83178919  
 A;Accession: S09174  
 A;Molecule type: mRNA  
 A;Residues: 623-742; 'A', 744-764, 'X', 766-827, 'A', 829-930, 'P', 832-836, 'P', 838-1097, 'L',  
 A;Cross-references: GB:J00113; GB:V00503; NID:930123; PIDN:CAA23761.1; PID:g825646

A; Experimental source: skin fibroblast cells  
 R; Rorlino, A.; Zolezzi, F.; Walli, M.; Pignatti, P.F.; Cetta, G.; Brunelli, P.C.; Mottes, Hum. Mol. Genet. 3, 2201-2205, 1994  
 A; Title: Severe (type III) osteogenesis imperfecta due to glycine substitutions in the collagen type I pro-alpha 1(I) chain  
 A; Reference number: 154365; MUID:95187161  
 A; Accession: I68653  
 A; Status: preliminary; translated from GB/EMBL/DDBJ  
 A; Molecule type: mRNA  
 A; Residues: 663-675; 'V', '677', 'P', '679-742', 'A', '744-746 <FOR>  
 A; Cross-references: GB:47668; NID:9109095; PIDN:AAB59577\_1; PID:9109096  
 R; Niyyibizi, C.; Bonadio, J.; Byers, P.H.; Eyre, D.R.  
 J. Biol. Chem. 267, 23112, 1992  
 A; Accession: I55369  
 A; Status: translated from GB/EMBL/DDBJ  
 A; Molecule type: mRNA  
 A; Residues: 665-666; 'D', '668-670 <WIV>  
 A; Cross-references: GB:J00613; NID:9180888; PIDN:AAB59384\_1; PID:9180889  
 A; Note: mutant sequence from a patient with osteogenesis imperfecta type I  
 R; Batteman, J.F.; Hannagan, M.; Chan, D.; Cole, W.G.  
 Biochem. J. 276, 757-770, 1991  
 A; Title: Characterization of a type I collagen alpha 2(I) glycine-586 to valine substitution method  
 A; Reference number: A56799; MUID:91291136  
 A; Accession: A56799  
 A; Molecule type: mRNA  
 A; Residues: 672-675; 'V', '677', 'P', '679-681 <BAT>  
 A; Cross-references: GB:S39873; NID:9169911; PIDN:AAB19314\_1; PID:9222761  
 A; Note: sequence extracted from NCBI backbone (NCBIN:19878; NCBIP:39186)  
 A; Note: mutant sequence of patient with osteogenesis imperfecta type IV; the authors suggest control sequence  
 R; Maekelie, J.K.; Vuorio, T.; Vuorio, E.  
 Biochim. Biophys. Acta 1049, 171-176, 1990  
 A; Title: Growth-dependent modulation of type I collagen production and mRNA levels in cultured fibroblasts  
 A; Reference number: S10768; MUID:90304220  
 A; Accession: S10768  
 A; Molecule type: mRNA  
 A; Residues: 960-1021; 'L', '1023-1188', 'D', '1190-1197', 'S', '1199-1356 <MAE>  
 A; Cross-references: EMBL:X5525; NID:910101; PIDN:CAA39142\_1; PID:93102  
 A; Experimental source: fibroblast cell culture  
 R; Myers, J.C.; Chu, M.L.; Faro, S.H.; Clark, W.J.; Prockop, D.J.; Ramirez, F.  
 Proc. Natl. Acad. Sci. U.S.A. 78, 3516-3520, 1981  
 A; Title: Cloning a cDNA for the pro-alpha2 chain of human type I collagen.  
 A; Reference number: A18855; MUID:91273090  
 A; Accession: A18855  
 A; Molecule type: mRNA  
 A; Residues: 964-979; 'V', '981-1018', 'Q', '1020 <WVE>  
 A; Cross-references: GB:J00114; NID:9180393; PIDN:AAA51996\_1; PID:9180394  
 A; Note: 1019-leu was also found  
 R; Wenstrup, R.J.; Cohn, D.H.; Cohen, T.; Byers, P.H.  
 J. Biol. Chem. 263, 734-7740, 1988  
 A; Title: Arginine for glycine substitution in the triple-helical domain of the products of the human type I collagen gene  
 A; Reference number: I55285; MUID:88222975  
 A; Accession: I55285  
 A; Status: translated from GB/EMBL/DDBJ  
 A; Molecule type: DNA  
 A; Residues: 1090-1107 <WN1>  
 A; Cross-references: GB:M22815; NID:917602; PIDN:AAA51844\_1; PID:917603  
 A; Status: translated from GB/EMBL/DDBJ  
 A; Molecule type: DNA  
 A; Residues: 1090-1101; 'R', '1103-1107 <WN2>  
 A; Cross-references: GB:M22817; NID:917606; PIDN:AAA51846\_1; PID:917607  
 A; Note: mutant sequence from a patient with osteogenesis imperfecta type IV  
 R; Myers, J.C.; Dickson, L.A.; de Wet, W.J.; Bernard, M.P.; Chu, M.L.; di Liberto, M.; Peijnenburg, J.; Biol. Chem. 258, 10128-10135, 1983  
 A; Title: Analysis of the 3' end of the human pro-alpha-2(I) collagen gene. Utilization of the 3' end of the gene  
 A; Reference number: S09175; MUID:83290053  
 A; Accession: S09175  
 A; Molecule type: DNA

R; Pariente, M.G.; Chung, L.C.; Ryynanen, J.; Woodley, D.T.; Wynn, K.C.; Bauer, Proc. Natl. Acad. Sci. U.S.A. 88: 6931-6935, 1991  
 A; Title: Human type VII collagen: cDNA cloning and chromosomal mapping of the gene.  
 A; Accession: SI6316  
 A; Molecule type: mRNA  
 A; Residues: 815-892, 'E', 894-1439 <PAR>  
 A; Cross-references: GB: M65158; S49017; NID: 9160914; PIDN: AAA96439.1; PID: 9180915  
 A; Experimental source: keratinocyte  
 R; Gammon, W.R.; Abernethy, M.L.; Padilla, K.M.; Prisayanh, P.S.; Cook, M.E.; Wright, J.; J. Invest. Dermatol. 99, 691-696, 1992  
 A; Title: Noncollagenous (NC1) domain of collagen VII resembles multidomain adhesion protein  
 A; Reference number: 156328; MUID: 93107742  
 A; Accession: I56328  
 A; Status: translated from GB/EMBL/DDBJ  
 A; Molecule type: mRNA  
 A; Residues: 'FPR', 372-517, 'DV', 520-540, 'W', 542-1255 <RES>  
 A; Cross-references: GB: S51235; NID: 9262308; PIDN: AAB2437.1; PID: 9262309  
 R; Seltzer, J.L.; Eisen, A.Z.; Bauer, E.A.; Morris, N.P.; Glanville, R.W.; Burgeson, R.E.; J. Biol. Chem. 264: 3822-3826, 1989  
 A; Title: Cleavage of type VII collagen by interstitial collagenase and type IV collagenase  
 A; Reference number: A30296; MUID: 89139437  
 A; Accession: A30296  
 A; Molecule type: protein  
 A; Residues: 'A', 1240-1246, 'G', 1248-1250, 'XE', 1253-1255, 'Q', 1257, 'E', 2032, 'C', 2034-2041; 'A'; Note: two reported peptides cannot be reliably located  
 Hum. Mol. Genet. 2, 273-278, 1993  
 A; Title: The carboxyl-terminal half of type VII collagen, including the non-collagenous R; Greenbaum, D.S.  
 A; Reference number: 148103; MUID: 9327185  
 A; Accession: I84686  
 A; Status: preliminary; translated from GB/EMBL/DDBJ  
 A; Molecule type: mRNA  
 A; Residues: 2395-2871, 'S', 2873-2944 <RES>  
 A; Cross-references: GB: L06862; NID: 9388713; PIDN: AAA89196.1; PID: 9388714  
 R; Chiricano, A.M.; Ryynaenen, M.; Utton, J.  
 Proc. Natl. Acad. Sci. U.S.A. 91: 3549-3553, 1994  
 A; Title: Dominant dystrophic epidermolysis bullosa: identification of a Gly --> Ser substitution  
 A; Reference number: A5255; MUID: 94224777  
 A; Contents: annotation  
 C; Comment: Prolines and lysines at the third position of the tripeptide repeating unit are ed and subsequently O-glycosylated.  
 C; genetics:  
 A; Gene: GDB:COL7A1; EBRI; EBI; EB  
 A; Cross-references: GDB:128755; OMIM:120120  
 A; Map position: 3p21.3-3p21.3  
 A; Note: defects in this gene can result in dominant and recessive dystrophic epidermolys  
 A; Note: there are 118 introns  
 C; Complex: type VII collagen is probably a homotrimer  
 C; Function:  
 A; Description: structural component of extracellular polymer associated with anchoring f  
 C; Superfamily: unassigned collagens; animal Kunzitz-type proteinase inhibitor homology; f  
 C; Keywords: coiled coil; extracellular matrix; glycoprotein; hydroxylysine; hydroxyprolin  
 F; 1-17/294/Product: collagen alpha 1(VII) chain #status predicted <SIC>  
 F; 17-123/Domain: amino-terminal nonhelical #status predicted <MAT>  
 F; 17-201/Domain: von Willebrand factor type A repeat homology <WMA>  
 F; 231-318/Domain: fibronectin type III repeat homology <FN1>  
 F; 327-413/Domain: fibronectin type III repeat homology <FN2>  
 F; 414-502/Domain: fibronectin type III repeat homology <FN3>  
 F; 508-593/Domain: fibronectin type III repeat homology <FN4>  
 F; 598-603/Domain: fibronectin type III repeat homology <FN5>  
 F; 686-771/Domain: fibronectin type III repeat homology <FN6>  
 F; 776-862/Domain: fibronectin type III repeat homology <FN7>  
 F; 864-952/Domain: fibronectin type III repeat homology <FN8>  
 F; 954-1045/Domain: fibronectin type III repeat homology <FN9>  
 F; 1052-1219/Domain: von Willebrand factor type A repeat homology <VWA2>  
 F; 1170-1172/Region: cell attachment (R-G-D) motif  
 F; 1189-1253/Region: cysteine/proline-rich  
 F; 1254-2783/Region: interrupted helical  
 F; 1334-1336/Region: cell attachment (R-G-D) motif  
 F; 2008-2010/Region: cell attachment (R-G-D) motif  
 F; 2553-2555/Region: cell attachment (R-G-D) motif

F;2784-2944/Domain: carboxyl-terminal nonhelical #status predicted <NC2>  
 F;2876-2929/Domain: animal Kinase-type proteinase inhibitor homology <BPI>  
 F;3377-786-1109/Binding site: carbohydrate (Asn) (covalent) #status predicted  
 F;2167-2176-2185-2188-2664-2667-2673/Modified site: 4-hydroxyproline (Pro) #status ex  
 F;2625-2631/Modified site: 5-hydroxylysine (LyS) #status experimental  
 F;2652-2654-2802-2804/Disulfide bonds: interchain #status predicted  
 F;2654-2802-2804/Disulfide bonds: interchain #status predicted  
 Query Match 65.2% Score 60; DB 2; Length 2944;  
 Best Local Similarity 66.7% Pred. No. 2.2; Mismatches 2;  
 Matches 10; Conservative 3; Indels 0; Gaps 0;  
 Qy 1 PAGPWGPNKGDKVQ 15  
 Db 1623 PPGPVGVGRDGEVG 1637

**RESULT 6**  
 S10847  
 collagen alpha 2(I) chain - chicken (fragment)  
 C;Species: Gallus gallus (chicken)  
 C;Date: 21-Nov-1993 #sequence\_revision 26-Jul-1996 #text\_change 23-May-1997  
 C;Accession: S10847; S65690  
 R;Dickson, L.A.; Ninomiya, Y.; Bernard, M.P.; Pesciotta, D.M.; Parsons, J.; Green, G.  
 J. Biol. Chem. 256, 8407-8415, 1981  
 A;Title: The exon/intron structure of the 3'-region of the pro-alpha-2(I) collagen gene  
 A;Reference number: S10847; MUID:81264246  
 A;Accession: S10847  
 A;Molecule type: DNA  
 A;Residues: 1-365 <DIX>  
 A;Cross-references: EMBL:J00811  
 A;Note: the authors translated the codon ATA for residue 207 as Asp, AGC for residue  
 A;Accession: S65690  
 A;Molecule type: Protein  
 A;Residues: 121-123, 'X', 125-134, 'X', 136-137, 'X', 139, 'XXX', 143 <DIX>  
 C;Genetics:  
 A;Gene: COLIA2  
 A;Introns: 36/3; 54/3; 90/3; 173/1; 236/3; 317/3  
 C;Superfamily: collagen alpha 2(I) chain; fibrillar collagen carboxyl-terminal homolog  
 C;Keywords: coiled coil; extracellular matrix; glycoprotein; trimer; triple helix  
 F;136-365/domain: fibrillar collagen carboxyl-terminal homology <FC>

Query Match 62.0% Score 57; DB 2; Length 365;  
 Best Local Similarity 66.7% Pred. No. 0.74; Mismatches 2; Indels 0; Gaps 0;  
 Matches 10; Conservative 2; MisMatches 3; Indels 0; Gaps 0;

Qy 1 PAGPWGPNKGDKVQ 15  
 Db 56 PHGPSPGHKGDKRNG 70



A;Residues: 1-636 <GLUD>  
A;Cross-references: EMBL:X70369; NID:957915; PIDN:CAA49832.1; PID:g57916  
R;Frankel, F.R.; Hsu, C.Y.J.; Myers, J.C.; Lin, E.; Lytle, C.R.; Komm, B.; Mohn, K.  
DNA 7, 347-354, 1988  
A;Title: Regulation of alpha-2 (I), alpha-1 (III), and alpha-2(V) collagen mRNAs by ester  
A;Reference number: A29905; MUID:88296083  
A;Accession: A29905  
A;Molecule type: mRNA  
A;Residues: 308-482 <FRRA>  
A;Cross-references: GB:M21354; NID:9203500; PIDN:AAA40942.1; PID:9203501  
R;Glunoff, V.; Maekela, J.K.; Vuorio, E.  
submitted to the EMBL Data Library, February 1993  
A;Accession number: S31924  
A;Status: preliminary  
A;Accession: S31924  
A;Molecule type: mRNA  
A;Residues: 2-636 <GL2>  
A;Cross-references: EMBL:X70369  
C;Keywords: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;  
C;Keywords: coiled coil; extracellular matrix; glycoprotein; trimer; triple helix  
F;408-636/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match      Best Local Similarity 60.9%; Score 56; DB 2; Length 1076;  
Best Local Similarity 66.7%; Pred. No. 3.1; Mismatches 4; Indels 0; Gaps 0;  
Matches 10; Conservative 1; Mismatches 1; Mismatches 4; Indels 0; Gaps 0;

Qy      1 PAGPWPNGKDGKVG 15  
Db      313 PVGPHGPKGSQG 327

RESULT 13

B40333 collagen alpha 1(II) chain precursor - African clawed frog  
C;Species: Xenopus laevis (African clawed frog)  
C;Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
C;Accession: B40333  
R;Su, M.W.; Suzuki, H.R.; Bieler, J.J.; Solursh, M.; Ramirez, F.  
J. Cell Biol. 115, 565-575, 1991  
A;Title: Expression of two nonallelic type II procollagen genes during Xenopus laevis  
A;Reference number: A40333; MUID:92011898  
A;Accession: B40333  
A;Status: preliminary  
A;Accession: B40333  
A;Molecule type: mRNA  
A;Residues: 1-1486 <UUA>  
A;Cross-references: GB:M3595  
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology <WMC>  
F;37-96/Domain: von Willebrand factor type C repeat homology <WMC>  
F;1258-1486/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match      Best Local Similarity 60.9%; Score 56; DB 1; Length 1486;  
Best Local Similarity 66.7%; Pred. No. 4.3; Mismatches 4; Indels 0; Gaps 0;  
Matches 10; Conservative 1; Mismatches 1; Mismatches 4; Indels 0; Gaps 0;

Qy      1 PAGPWPNGKDGKVG 15  
Db      1167 PPGPVGPGSKGSNG 1181

RESULT 14

A40333 collagen alpha 1'(II) chain precursor - African clawed frog  
C;Species: Xenopus laevis (African clawed frog)  
C;Date: 16-Sep-1992 #sequence\_revision 16-Sep-1992 #text\_change 16-Jul-1999  
C;Accession: A40333  
R;Su, M.W.; Suzuki, H.R.; Bieker, J.J.; Solursh, M.; Ramirez, F.  
J. Cell Biol. 115, 565-575, 1991  
A;Title: Expression of two nonallelic type II procollagen genes during Xenopus laevis  
A;Reference number: A40333; MUID:92011898  
A;Accession: A40333  
A;Status: nucleic acid sequence not shown  
A;Molecule type: mRNA  
A;Residues: 1-1492 <SUAA>  
A;Cross-references: GB:M3596  
A;Note: this sequence is presented as substitutions relative to another sequence in a  
es they replace; the appropriate interpretation of the sequence figure was reconstruc  
C;Superfamily: collagen alpha 1(II) chain; fibrillar collagen carboxyl-terminal homolo  
C;Keywords: coiled coil; extracellular matrix; glycoprotein; trimer; triple helix  
F;37-96/Domain: von Willebrand factor type C repeat homology <WMC>  
F;1263-1492/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match      Best Local Similarity 60.9%; Score 56; DB 2; Length 1492;  
Best Local Similarity 66.7%; Pred. No. 4.3; Mismatches 4; Indels 0; Gaps 0;  
Matches 10; Conservative 1; Mismatches 1; Mismatches 4; Indels 0; Gaps 0;

Qy      1 PAGPWPNGKDGKVG 15  
Db      1170 PPGPVGPGSKGSNG 1184

RESULT 15

B40333 collagen alpha 1(II) chain precursor - African clawed frog  
C;Species: Xenopus laevis (African clawed frog)  
C;Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
C;Accession: B40333  
R;Su, M.W.; Suzuki, H.R.; Bieler, J.J.; Solursh, M.; Ramirez, F.  
J. Cell Biol. 115, 565-575, 1991  
A;Title: Expression of two nonallelic type II procollagen genes during Xenopus laevis  
A;Reference number: A40333; MUID:92011898  
A;Accession: B40333  
A;Status: preliminary  
A;Accession: B40333  
A;Molecule type: mRNA  
A;Residues: 1-1486 <UUA>  
A;Cross-references: GB:M3595  
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homolo  
F;37-96/Domain: fibrillar collagen carboxyl-terminal homology <WMC>

RESULT 17

CGHUE  
 collagen alpha 1(XI) chain precursor - human  
 N; Alternate names: procollagen alpha 1(XI) chain  
 C; Date: 31-Mar-1990 #sequence\_revision 03-Oct-1995 #text\_change 08-May-1998  
 C; Accession: A35239; A31795  
 R; Yoshioka, H.; Ramirez, F.  
 J. Biol. Chem. 265, 6423-6426, 1990  
 A; Title: Pro-alpha1(XI) collagen. Structure of the amino-terminal propeptide and expression  
 A; Reference number: A35239; MUID:9020294  
 A; Accession: A35239  
 A; Molecular type: mRNA  
 A; Residues: 1-558 <XOS>  
 A; Cross-references: GB:J05407  
 R; Bernard, M.; Yoshioka, H.; Rodriguez, E.; van der Rest, M.; Kimura, T.; Ninomiya, Y.;  
 J. Biol. Chem. 261, 17159-17166, 1988  
 A; Title: Cloning and sequencing of pro-alpha1(XI) collagen cDNA demonstrates that type XI  
 A; Reference number: A92689; MUID:89034222  
 A; Accession: A31795  
 A; Molecular type: DNA; mRNA  
 A; Residues: 530-1806 <BERZ>  
 A; Cross-references: GB:J04177  
 A; Note: parts of this sequence were determined by protein sequencing  
 C; Comment: Prolines and lysines at the third position of the tripeptide repeating unit  
 ed and subsequently O-glycosylated.  
 C; Genetics:  
 A; Gene: GDB:COL1A1; COL6  
 A; Cross-references: GDB:120555; OMIM:120280  
 A; Map position: 1p21-1p21  
 A; Introns: 561/3; 579/3; 597/3; 615/3; 633/3; 648/3; 666/3; 681/3  
 A; Note: the list of introns is incomplete  
 C; Complex: type XI collagen may be a heterotrimer of two alpha 1(XI) chains and one alpha 3(XI) chain (see PIR:CGHUC6), initially linked by disulfide bonds among their carboxyl-terminal  
 rmed with desmosine cross-links made from lysine and allysine residues  
 C; Function:  
 A; Description: structural component of extracellular fibrous polymer associated with cell  
 A; Note: may play a role in controlling the lateral growth of collagen II fibrils  
 C; Superfamily: collagen alpha 1(V) chain; fibrillar collagen carboxy-terminal homology  
 C; Keywords: coiled coil; extracellular matrix; glycoprotein; hydroxylysine; hydroxyproline  
 F; 1-36/Domain: signal sequence #status predicted <SIG>  
 F; 35-260/Domain: PRRP-like #status predicted <PARP>  
 F; 37-511/Domain: amino-terminal propeptide #status predicted <PRO>  
 F; 512-1565/Product: collagen alpha 1(XI) chain #status predicted <MAT>  
 F; 512-57/Region: amino-terminal nonhelical telopeptide  
 F; 1543-1565/Region: carboxy-terminal nonhelical telopeptide  
 F; 1566-1806/Domain: carboxy-terminal propeptide #status predicted <CTP>  
 F; 1583-1805/Domain: fibrillar collagen carboxy-terminal homology <FCC>  
 F; 161-243-182-236/Disulfide bonds: #status predicted  
 F; 505/Modified site: allysine (Lys) #status predicted  
 F; 612,1457/Modified site: 5-hydroxylysine (Lys) #status predicted  
 F; 612,1452/Binding site: carbohydrate (Lys) (covalent) #status predicted

RESULT 18

Query Match 60.9%; Score 56; DB 1; Length 1806;  
 Best Local Similarity 66.7%; Pred. No. 5.2;  
 Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 Qy 1 PAGPMGPNGDKYVG 15  
 Db 1226 PRGPGPQGADGPOG 1240

RESULT 19

T16841  
 hypothetical protein T10E10.2 - *Caenorhabditis elegans*  
 C; Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 02-Jun-2000  
 C; Accession: T16841  
 R; Geisel, C.  
 submitted to the EMBL Data Library, October 1995  
 A; Description: The sequence of *C. elegans* cosmid T10E10.  
 A; Gene: Z1888  
 A; Accession: T16841  
 A; Status: preliminary; translated from GB/EMBL/DDJB  
 A; Residues: 1-326 <GEI>  
 A; Cross-references: EMBL:U39644; NID:91049339; PTD:91049341; PIDN:AAA80350.1; CESP:T10E10  
 C; Genetics:  
 A; Gene: CESP:T10E10.2  
 C; Superfamily: unassigned collagens  
 Query Match 59.8%; Score 55; DB 2; Length 326;  
 Best Local Similarity 66.7%; Pred. No. 1.3;  
 Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 Qy 1 PAGPMGPNGDKYVG 15  
 Db 167 PAGPKGPNGNSGDG 181

RESULT 20

T16831  
 hypothetical protein T07H6.3 - *Caenorhabditis elegans*  
 C; Species: *Caenorhabditis elegans*  
 C; Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 18-Feb-2000  
 C; Accession: T16831  
 R; Geisel, C.  
 submitted to the EMBL Data Library, April 1996  
 A; Description: The sequence of *C. elegans* cosmid T07H6.  
 A; Reference number: 218586  
 A; Accession: T16831  
 A; Status: preliminary; translated from GB/EMBL/DDJB  
 A; Molecular type: DNA  
 A; Residues: 1-363 <GEI>  
 A; Cross-references: EMBL:U53344; NID:91255886; PTD:91255887; PIDN:AAA96223.1; GSPDB:G  
 A; Experimental source: strain Bristol N2; clone T07H6  
 C; Genetics:  
 A; Gene: CESP:T07H6.3  
 A; Map position: X  
 A; Introns: 62/3  
 C; Superfamily: unassigned collagens  
 Query Match 59.8%; Score 55; DB 2; length 363;  
 Best Local Similarity 66.7%; Pred. No. 1.4;

Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
Qy 1 PAGPWGPNGKDGKVG 15  
| :| :| :| :| :| :| :| :| :|  
Db 229 PAGPKGPNGNSGSDG 243

**RESULT 21**

S59513 collagen II Al protein - zebra fish (fragment)  
C;Species: Brachydanio rerio (zebra fish)  
C;Date: 12-Feb-1998 #sequence\_revision 20-Feb-1998 #text\_change 13-Aug-1999  
C;Accession: S59513  
R;Yan, Y.; Hotta, K.; Riddleman, B.; Postlethwait, J.H.  
A;Description: Expression of a type II collagen gene in the zebrafish embryonic axis.  
A;Accession: S59513  
A;Status: preliminary  
A;Molecule type: mRNA  
A;Residues: 1-164 <PAN>  
A;Cross-references: EMBL:U23822; NID:973660; PIDN:AAA96815.1; PID:973661  
C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology; F;236-464>/domain: fibrillar collagen carboxyl-terminal homology <FCC>

Query Match 59.8%; Score 55; DB 2; Length 464;  
Best Local Similarity 66.7%; Pred. No. 1.9; Mismatches 1; Mismatches 4; Indels 0; Gaps 0;  
Qy 1 PAGPWGPNGKDGKVG 15  
| :| :| :| :| :| :| :| :| :|  
Db 142 PSGPVGPAGKDGSNG 156

**RESULT 22**

A05269 collagen alpha 1(III) chain precursor - chicken (fragments)  
C;Species: Gallus gallus (chicken)  
C;Date: 05-Jun-1987 #sequence\_revision 05-Jun-1987 #text\_change 16-Jul-1999  
C;Accession: A05269; A38035; A20855  
R;Yamada, Y.; Liaw, G.; Mudryj, M.; Obici, S.; de Crombrughe, B.  
Nature 310, 333-337, 1984  
A;Title: Conservation of the sizes for one but not another class of exons in two chick c  
A;Reference number: A05269; MUID:8427056  
A;Accession: A05269  
A;Molecule type: DNA  
A;Residues: 1-129 <PAM>  
A;Note: the authors translated the codon AAA for residue 68 as Arg, AAC for residue 173  
R;Yamada, Y.; Ruehn, K.; de Crombrughe, B.  
Nucleic Acids Res. 11, 2733-2744, 1983  
A;Reference number: A38035; MUID:83220816  
A;Accession: A38035  
A;Molecule type: DNA  
A;Residues: 330-615 <YAM2>  
A;Note: the authors translated the codon GAT for residue 548 as Glu  
R;Yamada, Y.; Mudryj, M.; Sullivan, M.; de Crombrughe, B.  
J. Biol. Chem. 258, 2758-2761, 1983  
A;Title: Isolation and characterization of a genomic clone encoding chick alpha1 type II  
A;Reference number: A20855; MUID:83135706  
A;Accession: A20855  
A;Molecule type: DNA  
A;Residues: 222-233, 'R', 235-236, 'V', 238-239/330-336, 'V', 338-351 <YAM>  
A;Note: the authors translated the codon CGT for residue 234 as GLY, GTT for residue 237  
C;Comment: Chicken collagen alpha 1(III) chain has about 50 exons. This sequence corresponds to the 240-275', 6(276-293), 5(294-329), 4(330-423), 3(424-484), 2(485-567), and 1(568-615) C;Genetics: introns: 68/3; 86/3; 119/3; 134/3; 152/3; 188/3; 221/3; 239/3; 275/3; 293/3; 423/3; 46 C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology; F;2-68/Region: amino-terminal propeptide nonhelical F;2-68/Domain: von Willebrand factor type C repeat homology <WVC> F;69-347>/domain: collagen alpha helical chain (fragments) #status predicted <CAH> F;348-615>/domain: carboxyl-terminal propeptide #status predicted <CTP>

**RESULT 23**

CGCH2S collagen alpha 2(I) chain Precursor - chicken (fragments)  
C;Species: Gallus gallus (chicken)  
C;Date: 24-Apr-1984 #sequence\_revision 15-Aug-1997 #text\_change 21-Jul-2000  
C;Accession: I50206; I50207; S07354; S10848; S10400; S11146; I50628; I50170; I50625; R;Boedtker, H.; Finer, M.; Ahn, S.  
Ann. N.Y. Acad. Sci. 460, 85-116, 1985  
A;Title: The structure of the chicken alpha 2 collagen gene.  
A;Reference number: I50206; MUID:86185168  
A;Accession: I50206  
A;Status: translated from GB/EMBL/DDJB  
A;Molecule type: DNA  
A;Residues: 1-245 <PAN>  
A;Cross-references: GB:M25963; NID:9211581; PIDN:AAA69960.1; PID:9211605  
A;Accession: I50207  
A;Status: translated from GB/EMBL/DDJB  
A;Molecule type: DNA  
A;Residues: 246-431 <BO2>  
A;Cross-references: GB:M25965; NID:9211583; PIDN:AAA69961.1; PID:9211606  
R;Aho, S.; Tate, V.; Boedtker, H.  
Nucleic Acids Res. 12, 6117-6125, 1984  
A;Title: Location of the 11 bp exon in the chicken pro alpha-2(I) collagen gene.  
A;Reference number: S07354; MUID:84297217  
A;Accession: S07354  
A;Molecule type: DNA  
A;Cross-references: EMBL:X00760; NID:963266; PIDN:CAA25330.1; PID:963267  
R;Rate, V.E.; Finer, M.H.; Boedtker, H.; Doty, P.  
Nucleic Acids Res. 11, 91-104, 1983  
A;Title: Chick pro-alpha-2(I) collagen gene: exon location and coding potential for  
A;Reference number: S10480; MUID:83246518  
A;Accession: S10480  
A;Molecule type: RNA  
A;Residues: 1-89 <PAN>  
A;Cross-references: EMBL:X02657; NID:963314; PIDN:CAA26493.1; PID:963315  
A;Molecule type: DNA  
A;Residues: 17-73 <PAN>  
A;Note: the authors translated the codon CAG for residue 42 as Glu  
R;Vogeli, G.; Ohkubo, H.; Sobel, M.E.; Yamada, Y.; Pastan, I.; de Crombrughe, B.  
Proc. Natl. Acad. Sci. U.S.A. 78, 5334-5338, 1981  
A;Title: Structure of the promoter for chicken alpha-2 type I collagen gene.  
A;Reference number: S11146; MUID:82060240  
A;Accession: S11146  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-15 <VOC>  
A;Cross-references: EMBL:J00821  
R;Yamada, Y.; Avvedimento, V.E.; Mudryj, M.; Ohkubo, H.; Vogeli, G.; Irani, M.; Pasta Cell 22, 887-892, 1980  
A;Title: The collagen gene: evidence for its evolutionary assembly by amplification of  
A;Reference number: I50170; MUID:8112157  
A;Accession: I50170  
A;Status: preliminary; translated from GB/EMBL/DDJB  
A;Molecule type: DNA  
A;Residues: 126-161 <PAN>  
A;Cross-references: EMBL:V00400; NID:963305; PID:9833611  
A;Accession: I50170  
A;Status: preliminary; translated from GB/EMBL/DDJB

F;387-615>/domain: fibrillar collagen carboxyl-terminal homology <FCC>  
F;347-348/Disulfide bonds: interchain #status experimental  
F;516/Binding site: carbohydrate (Asn) #status covalent #status experimental

**RESULT 24**

Qy 1 PAGPWGPNGKDGKVG 15  
| :| :| :| :| :| :| :| :| :|  
Db 295 PPGPAGPPGKDGRRGG 309

Query Match 59.8%; Score 55; DB 2; Length 615;  
Best Local Similarity 66.7%; Pred. No. 2.5; Mismatches 1; Indels 4; Gaps 0; Gaps 0;  
Matches 10; Conservative 66.7%; Pred. No. 2.5; Mismatches 1; Indels 0; Gaps 0;

**RESULT 25**

Qy 1 PAGPWGPNGKDGKVG 15  
| :| :| :| :| :| :| :| :| :|  
Db 295 PPGPAGPPGKDGRRGG 309

Query Match 59.8%; Score 55; DB 2; Length 615;  
Best Local Similarity 66.7%; Pred. No. 2.5; Mismatches 1; Indels 4; Gaps 0; Gaps 0;

A; Molecule type: DNA  
A; Residues: 126-161 <Y12>  
A; Cross-references: GB:J00828; NID:9211295; PID:AAA51612.1; PID:9211317  
A; Status: preliminary; translated from GB/EMBL/DDBJ  
A; Molecule type: DNA  
A; Residues: 530-558 <Y3>  
A; Cross-references: EMBL:V00396; NID:963295; PID:9833609  
A; Accession: 150626  
A; Status: preliminary; translated from GB/EMBL/DDBJ  
A; Molecule type: DNA  
A; Residues: 572-601 <Y4>  
A; Cross-references: EMBL:V00398; NID:963299; PID:9833610  
A; Accession: 150624  
A; Status: preliminary; translated from GB/EMBL/DDBJ  
A; Molecule type: DNA  
A; Residues: 433-433 <Y45>  
A; Cross-references: EMBL:V00394; NID:963287; PID:9833608  
R; Kang, A.H.; Gross, J.  
Biochemistry 9, 796-804, 1970  
A; Title: Amino acid sequence of cyanogen bromide peptides from the amino-terminal region  
A; Reference number: A90568; MUID:70131186  
A; Accession: A90568  
A; Molecule type: protein  
A; Residues: 2 ,79-92 <KAN>  
A; Experimental source: skin  
R; Kang, A.H.; Igashii, S.; Gross, J.  
Biochemistry 8, 3200-3204, 1969  
A; Title: Characterization of the cyanogen bromide peptides from the alpha<sub>2</sub> chain of chick  
A; Reference number: A90557; MUID:69205369  
A; Accession: A90557  
A; Molecule type: protein  
A; Residues: 2 ,79-95;398-409, 'A', 411, 'V', 413-428 <KN2>  
A; Experimental source: skin  
A; Note: the compositions of the six CNBr peptides were determined. CNBr0 is residues 93-  
R; Higlberger, J.H.; Kang, A.H.; Gross, J.  
Biochemistry 10, 610-616, 1971  
A; Title: Comparative studies on the amino acid sequence of the alpha<sub>2</sub>-CB2 peptides from  
A; Reference number: A90358; MUID:71115216  
A; Accession: B90358  
A; Molecule type: protein  
A; Experimental source: skin  
R; Lane, J.M.; Miller, E.J.  
Biochemistry 8, 2134-2139, 1969  
A; Title: Isolation and characterization of the peptides derived from the alpha<sub>2</sub> chain of  
A; Reference number: A90555; MUID:69206882  
A; Accession: A90555  
A; Molecule type: protein  
A; Residues: '2 ,79-95;398-409, 'A', 411, 'V', 413-428 <LAN>  
A; Note: the compositions of the six CNBr peptides were determined with those from skin a  
R; Igashii, S.; Kang, A.H.; Gross, J.  
Biochem. Biophys. Res. Commun. 38, 697-702, 1970  
A; Title: Renaturation and ordering by electron microscopy of the cyanogen bromide peptide  
A; Reference number: A90168; MUID:7018851  
A; Contents: annotation; skin; order of CNBr peptides  
R; Vuust, J.; Lane, J.M.; Fiedzak, P.P.; Miller, E.J.; Piez, K.A.  
Biochem. Biophys. Res. Commun. 38, 703-708, 1970  
A; Title: The order of the CNBr peptides from the alpha<sub>2</sub> chain of collagen.  
A; Reference number: A90169; MUID:70181852  
A; Contents: annotation; bone; order of CNBr peptides  
R; Mooney, J.; Hanahan, D.; Tate, V.; Boedeker, H.; Doty, P.  
Nature 294, 129-135, 1981  
A; Title: Structure of the pro alpha-2(I) collagen gene.  
A; Reference number: S0327; MUID:82050801  
A; Accession: S07327  
A; Molecule type: DNA  
A; Residues: 94-9213, 247-431;432, 693-774 <R022>  
A; Cross-references: EMBL:J00826  
R; Fuller, F.; Boedeker, H.  
Biochemistry 20, 996-1006, 1981

A; Title: Sequence determination and analysis of the 3' region of chicken pro-alpha 1 (I)  
A; Reference number: 150623; MUID:81160715  
A; Accession: 150623  
A; Status: preliminary; translated from GB/EMBL/DDBJ  
A; Molecule type: mRNA  
A; Residues: 506-830 'T', 831-903, 'N', 904-964 <FUL>  
A; Cross-references: EMBL:V00390; NID:963248; PID:CAA23688.1; PID:963249  
R; Avendamento, E.V.; Vogeli, G.; Yamada, Y.; Maizel, J.V.  
Cell 21, 689-696, 1980  
A; Title: Correlation between splicing sites within an intron and their sequence compl  
A; Reference number: 150172; MUID:81064671  
A; Status: preliminary; translated from GB/EMBL/DDBJ  
A; Molecule type: DNA  
A; Residues: 536-558 <AVV>  
A; Cross-references: GB:MI0581; NID:9211323; PID:AAA48637.1; PID:9211326  
R; Lehbrach, H.; Frischauft, A.M.; Hannan, D.; Woynay, J.; Fuller, F.; Crkvenjakov, R.;  
Proc. Natl. Acad. Sci. U.S.A. 75, 47-542, 1978  
A; Title: Construction and characterization of a 2.5-kilobase procollagen clone.  
A; Reference number: 150171; MUID:79074829  
A; Accession: 150171  
A; Status: preliminary; translated from GB/EMBL/DDBJ  
A; Molecule type: mRNA  
A; Residues: 484-505 <LEH>  
A; Cross-references: GB:J00837; NID:94530617; PID:AAA51614.1; PID:9211320  
C; Genetics:  
R; Kang, A.; Gene: COLA2  
A; Introns: 24/1; 27/3; 33/3; 45/3; 73/3; 92/3; 107/3; 125/3; 143/3; 161/3; 179/3; 197/  
A; Note: the list of introns is incomplete  
C; Superfamily: collagen alpha 2(I) chain; fibrillar collagen carboxyl-terminal homolo  
C; Keywords: blocked amino end; coiled coil; extracellular matrix; glycoprotein; hydroo  
F; 1-22/Domain: signal sequence #status predicted <SIG>  
F; 23-77/Domain: amino-terminal propeptide #status predicted <PRO>  
F; 78-964/Product: collagen alpha 2(I) chain (fragments) #status predicted <MATN>  
F; 737-894/Region: amino-terminal nonhelical telopeptide  
F; 737-954/Domain: fibrillar collagen carboxyl-terminal homology <FCC>  
F; 78/Modified site: blocked amino end (Gin) (in mature form) (probably pyrrolidone car  
F; 83/Modified site: allysine (Lys) #status experimental  
F; 224-245/Modified site: 4-hydroxyproline (Pro) #status experimental  
F; 866/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 59 8% ; score 55; DB 1; Length 964;  
Best Local Similarity 66.7%; Pre. No. 3.9;  
Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy	1 PAGGWGPNGDKGV 15
Db	657 PPGFSSGGPKDGKNG 671

**RESULT 24**  
CGCHHS  
collagen alpha 1(I) chain - chicken (tentative sequence) (fragments)  
C;Species: Gallus gallus (chicken)  
C;Date: 12-Aug-1981 #sequence\_revision 06-Jul-1982 #text\_change 31-Mar-2000  
C;Accession: A90458; A90181; A02857  
R; Higlberger, J.H.; Cobbett, C.; Dixit, S.N.; Yu, W.; Seyer, J.M.; Kang, A.H.; Gross,  
Biochemistry 21, 2048-2055, 1982  
A;Title: Amino acid sequence of chick skin collagen alpha1(I)-GBB and the complete pri  
A;Reference number: A90458; MUID:82231995  
A;Accession: A90458  
A;Molecule type: protein  
A;Residues: 1-1036 <HIG>  
A;Experimental source: skin  
A;Note: this is the latest in a series of papers from these workers elucidating the s  
R; Eyer, D.R.; Glimcher, M.J.  
Biochem. Biophys. Res. Commun. 48, 720-726, 1972  
A;Title: Evidence for a previously undetected sequence at the carboxyterminus of the  
A;Reference number: A90181  
A;Molecule type: protein  
A;Residues: 1037-1042 <BR>  
A;Experimental source: skin

A;Note: residues 1037-1042 above correspond to the carboxyl end of the protein  
 C;Comment: Lysines at positions 103, 700, 934, and 946 above may be hydroxylated in some  
 C;Comment: Most of the prolines at the third position of the tripeptide repeating unit  
 C;Comment: Pro-1002 is the only 3-hydroxyproline and the only hydroxylated proline in pc  
 C;Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;  
 C;Keywords: coiled coil; extracellular matrix; glycoprotein; pyroglutamic acid; trimer;  
 F;1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental

Query Match	59.8%	Score 55;	DB 1;	Length 1042;		
Best Local Similarity	66.7%	Pred. No.	4.2;			
Matches	10;	Conservative	1;	Mismatches		
Qy	1 PAGPWNPGNGDKGVG 15	4;	Indels	0;	Gaps	0;
Db	60 PAGPPCKNGDDGEAG 74					

RESULT 25  
 T45467  
 collagen alpha 1(II) chain precursor [imported] - horse  
 N;Alternate names: type II collagen  
 C;Species: Equus caballus (domestic horse)  
 C;Date: 31-Jan-2000 #sequence\_revision 31-Jan-2000  
 C;Accession: T45467  
 R;Richardson, D.W.; Dodge, G.R.  
 R;Submitted to the EMBL Data Library, June 1996  
 A;Description: Cloning of equine type II collagen and modulation of its expression in eq  
 A;Reference number: 222977  
 A;Accession: T45467  
 A;Status: preliminary; translated from GB/EMBL/DDBJ  
 A;Molecule type: mRNA  
 A;Residues: 1-1418 <RIC>  
 A;Cross-references: EMBL:U62528; PIDN:AA05773.1  
 C;Superfamily: collagen alpha 1(II) chain; fibrillar collagen carboxyl-terminal homology;  
 Query Match
 59.8% | Score 55; | DB 2; | Length 1464; || Best Local Similarity | 66.7% | Pred. No. | 5.9; |  |
Matches	10;	Conservative	0;	Mismatches	5;	Indels	0;	Gaps	0;
Qy	1 PAGPWNPGNGDKGVG 15	4;	Indels	0;	Gaps	0;			
Db	1096 PPGPPGPGSKGDGANG 110								

RESULT 26  
 S59856  
 collagen alpha 1(III) chain precursor - mouse  
 C;Species: Mus musculus (house mouse)  
 C;Accession: S59856; S62120; S16373  
 R;Toman, P.D.; de Crombrugge, B.  
 Gene 147, 161-168, 1994  
 A;Title: The mouse type-III procollagen-encoding gene: genomic cloning and complete DNA  
 A;Reference number: S59856; MUID:95011609  
 A;Accession: S59856  
 A;Molecule type: DNA  
 A;Residues: 1-1464 <TOM>  
 A;Cross-references: EMBL:X52046  
 R;Toman, P. D.  
 submitted to the EMBL Data Library, November 1994  
 A;Reference number: S62120  
 A;Accession: S62120  
 A;Molecule type: DNA  
 A;Residues: 1-866; 'G', 868-1464 <TOA>  
 A;Cross-references: EMBL:X52046; MUID:9575321; PIDN:CAA36279.1; PID:9575322  
 R;Metsaeranta, M.; Toman, D.; de Crombrugge, B.; Vuorio, E.  
 Biochim. Biophys. Acta 1089, 241-243, 1991  
 A;Title: Specific hybridization probes for mouse type I, II, III and IX collagen mRNAs.  
 A;Reference number: S16176; MUID:91274355  
 A;Accession: S16373  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1442-1464 <MET>

Query Match	59.8%	Score 55;	DB 1;	Length 1042;		
Best Local Similarity	66.7%	Pred. No.	4.2;			
Matches	10;	Conservative	1;	Mismatches		
Qy	1 PAGPWNPGNGDKGVG 15	4;	Indels	0;	Gaps	0;
Db	60 PAGPPCKNGDDGEAG 74					

RESULT 27  
 CGHUGC  
 collagen alpha 1(III) chain precursor [validated] - human  
 N;Alternate names: Procollagen alpha 1(III) chain  
 N;Contains: chondrocalcin; collagen alpha 1(II) chain precursor splice form 1; collag  
 C;Species: Homo sapiens (man)  
 C;Date: 28-May-1986 #sequence\_revision 01-Sep-1995 #text\_change 08-Dec-2000  
 C;Accession: A38513; S06715; S24270; A24828; S06466; A35428; A30147; A33116; S64674;  
 C;Accession: A38513; S06715; S24270; A24828; S06466; A35428; A30147; A33116; S64674;  
 C;Accession: A38513; S06715; S24270; A24828; S06466; A35428; A30147; A33116; S64674;  
 R;Ryan, M.C.; Sierski, M.; Sandell, L.J.  
 Genomics 8, 41-48, 1990  
 R;Su, M.W.; Lee, B.; Ramirez, F.; Machado, M.; Horton, W.  
 A;Title: The human type II procollagen gene: identification of an additional protein-  
 A;Reference number: A38513; MUID:91184811  
 A;Accession: A38513  
 A;Molecule type: DNA  
 A;Residues: 1-103 <RYA>  
 A;Cross-references: GB:MG0299; NID:9108083; PIDN:AAA73873.1; PID:9108084  
 R;Su, M.W.; Lee, B.; Ramirez, F.; Machado, M.; Horton, W.  
 A;Title: Nucleotide sequence of the full length cDNA encoding for human type II proco  
 A;Reference number: S06715; MUID:90067946  
 A;Accession: S06715  
 A;Molecule type: mRNA  
 A;Residues: 1-28, R, '99-1487 <SU2>  
 A;Cross-references: EMBL:X1648; NID:929515; PIDN:CAA34488.1; PID:929516  
 A;Note: alternative splice form 1  
 R;Vikkula, M.; Metsaeranta, M.; Syraenen, A.C.; Ala-Kokko, L.; Vuorio, E.; Peltonen,  
 Biochem. J. 285, 287-294, 1992  
 A;Title: Structural analysis of the regulatory elements of the type-II procollagen ge  
 A;Reference number: S24270; MUID:9334585  
 A;Accession: S24270  
 A;Status: translation not shown  
 A;Molecule type: DNA  
 A;Residues: 1-28 <VRK>  
 A;Cross-references: EMBL:X58709; GB:S40537; NID:935659  
 A;Note: this translation is not annotated in GenBank entry HSPROCOL, release 11.0  
 R;Nunez, A.M.; Kohno, K.; Martin, G.R.; Yamada, Y.  
 Gene 44, 11-16, 1985  
 A;Title: Promoter region of the human pro-alpha-1-(II)-collagen gene.  
 A;Reference number: A24828; MUID:87031574  
 A;Accession: A24828  
 A;Molecule type: DNA  
 A;Residues: 1-8; 'T', 10-28 <NUIN>  
 A;Cross-references: GB:W25698; NID:9108072; PIDN:AAA52051.1; PID:9553237  
 R;Baldwin, C.T.; Regnato, A.M.; Smith, C.; Jimenez, S.A.; Prockop, D.J.  
 Biochem. J. 262, 525-528, 1989  
 A;Title: Structure of cDNA clones coding for human type II procollagen  
 A;Reference number: S06496; MUID:90026318  
 A;Accession: S06496  
 A;Molecule type: mRNA  
 A;Residues: 7-28, 'R', 99-157, 'P', 159-440, 'G', 442-456, 'E', 458-640, 'A', 642-831, 'PA', 834,  
 A;Cross-references: EMBL:X16711; NID:930040; PIDN:CAA34683.1; PID:930041

- A; Note: alternative splice form 1  
 R; Ryan, M.C.; Sandell, L.J.  
 J. Biol. Chem. 265, 10344-10339, 1990  
 A; Title: Differential expression of a cysteine-rich domain in the amino-terminal propeptidase  
 A; Reference number: A35428; MUID:90285153  
 A; Accession: A35428  
 A; Status: not compared with conceptual translation  
 A; Molecule type: mRNA  
 A; Residues: 27-81, 'L', 83-103 <RYV2>  
 A; Note: alternative splice form 2; splicing appears to be under developmental regulation  
 R; Su, M.W.; Benson-Chanda, V.; Vissing, H.; Ramirez, F.  
 A; Reference number: A30147; MUID:89233138  
 A; Molecule type: DNA  
 A; Residues: 104-157, 'P', 159-236 <SUM>  
 A; Cross-references: GB:J03055; GB:M23660; GB:M25655; GB:M25656; GB:M25730; GB:M32168; GB:R; Ala-Kokko, L.; Baldwin, C.T.; Moskowitz, R.W.; Prockop, D.J.  
 Proc. Natl. Acad. Sci. U.S.A. 87, 6563-6566, 1990  
 A; Title: Single base mutation in the type II procollagen gene (COL2A1) as a cause of pri  
 A; Reference number: A91227; MUID:90370826  
 A; Accession: A33116  
 A; Molecule type: DNA  
 A; Residues: 171-172, 'C', 174-175 <ALA>  
 A; Note: mutant sequence from a family with family with primary generalized osteoarthritis  
 R; Diab, M.; Wu, J.J.; Eyre, D.R.  
 B1chem. J. 314, 327-332, 1996  
 A; Title: Collagen type IX from human cartilage: a structural profile of intermolecular c  
 A; Reference number: S6673; MUID:96195147  
 A; Accession: S6674  
 A; Molecule type: protein  
 A; Residues: 188-189, 'X', 191-195; 1224-1230, 'X', 1232-1236 <DXA>  
 R; Franc, S.; Marzin, E.; Boutillon, M.M.; Lafont, R.; Lechene de la Porte, P.; Herbage,  
 Eur. J. Biochem. 234, 125-131, 1995  
 A; Title: Immunohistochemical and biochemical analyses of 20000-25000-year-old fossil car  
 A; Reference number: S66514; MUID:96096730  
 A; Accession: S66514  
 A; Molecule type: protein  
 A; Residues: 243-261; 575-590, 756-763, 'X', 765-779 <FRAS>  
 R; Tiller, G.E.; Weis, M.A.; Polunho, P.A.; Gruber, H.E.; Rimoin, D.L.; Cohn, D.H.; Eyr  
 Am. J. Hum. Genet. 56, 388-395, 1995  
 A; Title: An RNA-splicing mutation (G>IVS20) in the type II collagen gene (COL2A1) in a  
 A; Reference number: 136867; MUID:95150028  
 A; Accession: 138867  
 A; Status: preliminary; translated from GB/EMBL/DBJ  
 A; Molecule type: DNA  
 A; Residues: 440, 'G', 442-456, 'E', 458-460, 'P', 482-509 <TTIL>  
 R; Ramirez, F.  
 A; Cross-references: EMBL:X13783; NID:9557033; PIDN:AMB60370.1; PID:9557054  
 R; Vakkula, M.; Peitonen, L.  
 FEBS Lett. 250, 171-174, 1989  
 A; Title: Structural analyses of the polymorphic area in type II collagen gene.  
 A; Reference number: S04892  
 A; Molecule type: mRNA  
 A; Residues: 501-676, 'A', 678-783, 'A', 785-831, 'PA', 834, 'F', 836-1214 <GRAM>  
 A; Cross-references: EMBL:X13783; NID:930037; PIDN:CAA32030.1; PID:930050  
 R; Vakkula, M.; Peitonen, L.  
 FEBS Lett. 250, 171-174, 1989  
 A; Title: Structural analyses of the polymorphic area in type II collagen gene.  
 A; Reference number: S04892  
 A; Accession: S05000; MUID:89323561  
 A; Molecule type: DNA  
 A; Residues: 530-640, 'A', 642-785 <VIK2>  
 A; Cross-references: EMBL:X15158; NID:929951; PIDN:CAA34278.1; PID:91335018; PIDN:CAA3427  
 PIDN:CAA34283.1; PID:91335023; PIDN:CAA34284.1; PID:91335024  
 R; Bogaert, R.; Tiller, G.E.; Weis, M.A.; Gruber, H.E.; Rimoin, D.L.; Cohn, D.H.; Eyr  
 J. Biol. Chem. 267, 22522-22526, 1992  
 A; Title: An amino acid substitution (Gly853-->Glu) in the collagen alpha 1(II) chain pro  
 A; Reference number: A44309; MUID:93054548  
 A; Status: nucleic acid sequence not shown; not compared with conceptual translation  
 A; Molecule type: DNA; mRNA  
 A; Residues: 752-831, 'PA', 834, 'F', 836-1005, 'K', 1007-1036, 'Q', 1038-1052, 'E', 1054-1068, 'T',  
 A; Cross-references: GB:L00977; NID:9180812; PIDN:AB23914.1; PID:9258774  
 A; Note: sequence extracted from NCBI backbone (NCBIP11723); parts of this sequence  
 J. Biol. Chem. 265, 10344-10339, 1990  
 A; Note: this translation is not annotated with perinatal lethal hypochondrogenesis  
 A; Note: mutant sequence associated with perinatal lethal hypochondrogenesis  
 R; Miller, G.E.; Rimoin, D.L.; Murray, L.W.; Cohn, D.H.  
 Proc. Natl. Acad. Sci. U.S.A. 87, 389-3893, 1990  
 A; Title: Tandem duplication within a type II collagen gene (COL2A1) exon in an indivi  
 A; Reference number: S16502; MUID:90251662  
 A; Molecule type: DNA  
 A; Residues: 1164-1184, 'GPSKGKANGAISPGPI', 1185-1199 <NTL2>  
 A; Cross-references: EMBL:M37126; NID:9180808; PIDN:AA52037.1; PID:9180809  
 A; Note: mutant sequence from a patient with spondyloepiphyseal dysplasia  
 R; Cheah, K.S.B.; Stoker, N.G.; Griffin, J.R.; Grosveld, F.G.; Solomon, E.  
 Proc. Natl. Acad. Sci. U.S.A. 82, 2555-2559, 1985  
 A; Title: Identification and characterization of the human type II collagen gene (COL2  
 A; Reference number: A02858; MUID:85190534  
 A; Accession: A02858  
 A; Molecule type: DNA  
 A; Residues: 1032-1056, 'N', 1058-1068, 'T', 1070-1487 <CHE>  
 A; Cross-references: GB:J00116; NID:9180395; PIDN:AA51997.1; PID:9180396  
 R; Elima, K.; Vuorio, T.; Vuorio, E.  
 Nucleic Acids Res. 15, 9499-9504, 1987  
 A; Title: Determination of the single polyadenylation site of the human pro-alpha-1(III  
 A; Reference number: A27280; MUID:88067771  
 A; Accession: A27280  
 A; Molecule type: DNA; mRNA  
 A; Residues: 1175-1487 <ELI>  
 A; Cross-references: EMBL:X06268; NID:930096; PIDN:CAA29604.1; PID:930097  
 A; Experimental source: fetal epiphyseal cartilage  
 R; van der Rest, M.; Rosenberg, L.C.; Olsen, B.R.; Poole, A.R.  
 Biochem. J. 237, 923-925, 1986  
 A; Title: Chondrocalcin is identical with the C-propeptide of type II procollagen.  
 A; Reference number: A57033; MUID:87099927  
 A; Accession: A57033  
 A; Molecule type: protein  
 A; Residues: 'XE', 1244-1245, 'N', 1248, 'X', 1250-1265; 1295-1305; 1395-1408 <VAN>  
 A; Note: Chondrocalcin identified as released collagen I(II) chain carboxyl-terminal p  
 R; Strom, C.M.; Upholt, W.B.  
 Nucleic Acids Res. 12, 1025-1038, 1984  
 A; Title: Isolation and characterization of genomic clones corresponding to the human  
 A; Reference number: A21733; MUID:84118798  
 A; Accession: A21733  
 A; Molecule type: DNA  
 A; Residues: 1245-1295 <STR1>  
 A; Cross-references: EMBL:X00339; EMBL:X00298; NID:9394699; PIDN:CAA25092.1; PID:94378  
 A; Accession: B21733  
 A; Molecule type: DNA  
 A; Residues: 894-909, 'PF', <STR2>  
 A; Cross-references: GB:K0185; NID:930035; PIDN:CAA25082.1; PID:91335032  
 R; Nunez, A.M.; Francomme, C.; Young, M.F.; Martin, G.R.; Yamada, Y.  
 Biochemistry 24, 6343-6348, 1985  
 A; Title: Isolation and partial characterization of genomic clones coding for a human  
 gene.  
 A; Reference number: A24561; MUID:86104139  
 A; Accession: A24561  
 A; Molecule type: DNA  
 A; Residues: 1296-1358 <NNUN2>  
 A; Cross-references: GB:W1208; NID:9180017  
 A; Note: this translation is not annotated in GenBank entry HUMCCT2A, release 11.0  
 A; Note: the codons given for 133-Lys (AGG) and 1350-Gly (GCA) are inconsistent with  
 R; Sangiorgi, F.O.; Benson-Chanda, V.; de Wet, W.J.; Sobel, M.E.; Tsipouras, P.; Ramir  
 Nucleic Acids Res. 13, 2207-2225, 1985  
 A; Title: Isolation and partial characterization of the entire human pro alpha 1(II) c  
 A; Reference number: I37249; MUID:85215609  
 A; Accession: S59491  
 A; Molecule type: DNA  
 A; Residues: 7-28 <SAN2>  
 A; Cross-references: GB:M23759; NID:9180845; EMBL:X03320; GB:M24938; NID:930104

A;Note: the GenBank PID is based on an incorrect reading frame  
A;Accession: I17250  
A;Status: translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 541-560 <SAN3>  
A;Cross-references: EMBL:X02378; GB:M23870; NID:930107; PIDN:CAA26227.1; PID:929621  
A;Accession: I17251

Query Match    59.8%;    Score 55;    DB 1;    Length 1487;  
Best Local Similarity    66.7%;    Pred. No. 6;    Mismatches 10;    Conservative 1;    Indels 4;    Gaps 0;  
Matches 10;    Mismatches 0;    Gaps 0;  
Db    1165 PPGPVGPSPKGDGANG 1179

RESULT 28

CGH2E  
collagen alpha 2(XI) chain precursor - human (fragment)  
N;Alternate names: procollagen alpha 2(XI) chain  
N;Contains: proline/arginine-rich protein (PARP)  
C;Species: Homo sapiens (man)  
C;Date: 07-Jun-1990 #sequence\_revision 03-Oct-1995 #text\_change 22-Jun-1999  
C;Accession: S34790; A32645  
R;Zhdikova, N.I.; Brewton, R.G.; Mayne, R.  
FEBs Lett. 326, 25-28, 1993  
A;Title: Molecular cloning of PARP (proline/arginine-rich protein) from human cartilage  
A;Reference number: S34790; MUID:9314796  
A;Accession: S34790  
A;Molecule type: mRNA  
A;Residues: 1-663 <ZHI>  
A;Cross-references: EMBL:118987; NID:9306439; PIDN:AA35498.1; PID:9306440  
R;Kimura, T.; Cheah, K.S.E.; Chan, S.D.H.; Lui, V.C.H.; Mattei, M.G.; van der Rest, M.;  
J. Biol. Chem. 264, 13910-13916, 1989  
A;Title: The human alpha2(XI) collagen (COL11A2) chain. Molecular cloning of cDNA and ge  
A;Reference number: A32645; MUID:9340485  
A;Molecule type: DNA; mRNA  
A;Residues: 586-1546 <KIM>  
A;Cross-references: GB:J04974; NID:9180714; PIDN:AAA52034.1; PID:9180715  
A;Note: parts of this sequence were determined by protein sequencing  
C;Comment: Prolines and lysines at the third position of the tripeptide repeating unit  
ed and subsequently O-glycosylated.  
C;Genetics:  
A;Gene: GDB:COL11A2  
A;Cross-references: GDB:119788; OMIM:120290  
A;Map position: 6p21.3-6p21.3  
A;Introns: 1302/3; 1322/3; 1350/3; 1440/1; 1477/3  
A;Note: the list of introns is incomplete  
C;Complex: type XI collagen may be a heterotrimer of two alpha 1(XI) chains (see PIR:CGH  
3(XI)) chain (see PIR:CGH06C), initially linked by disulfide bonds among their carboxyl-  
rmed with desmosine cross-links made from lysine and allysine residues  
C;Function:  
A;Description: structural component of extracellular fibrous polymer associated with cell  
A;Note: may play a role in controlling the lateral growth of collagen II fibrils  
C;Superfamily: collagen alpha 1(V) chain; fibrillar collagen carboxyl-terminal homology  
C;Keywords: coiled coil; extracellular matrix; glycoprotein; hydroxylysine; hydroxyproline  
F:1-254/Domain: non-collagenous (fragment) #status predicted <NC3>  
F:1-187/Product: proline/arginine-rich helix PARP protein (fragment) #status predicted <PARP>  
F:225-305/Domain: triple helix #status predicted <COL2>  
F:305-342/Domain: non-collagenous #status predicted <NC2>  
F:343-1356/Region: helical  
F:424-431/Region: cell attachment (R-G-D) motif  
F:447-449/Region: cell attachment (R-G-D) motif  
F:1125-1259/Region: cell attachment (R-G-D) motif  
F:1357-1380/Region: carboxyl-terminal nonhelical telopeptide  
F:1381-1546/Domain: carboxyl terminal propeptide (fragment) #status predicted <CTP>  
F:14403-1546/Domain: fibrillar collagen carboxyl-terminal homology (fragment) #status aty  
F:109-163 151-1545/Disulfide bonds: #status predicted  
F:319/Modified site: allysine (Lys) #status predicted  
F:426-1266/Modified site: 5-hydroxylysine (Lys) #status predicted

F:426-1266/Binding site: carbohydrate (Lys) (covalent) #status predicted  
F:927-933 1008-1017 1035-1130 1290-1296 1305-1317 1320/Modified site: 4-hydroxyprolin  
F:929/Modified site: 4-hydroxyproline (Pro) #status atypical  
F:942-1023 1299/Modified site: 5-hydroxylysine (Lys) #status experimental  
F:942-11023 1299/Binding site: carbohydrate (Lys) (covalent) #status experimental  
F:1427-1433 1450-1459/Disulfide bonds: interchain #status predicted  
F:1450/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match    59.8%;    Score 55;    DB 1;    Length 1546;  
Best Local Similarity    66.7%;    Pred. No. 6-3;  
Matches 10;    Conservative 0;    Mismatches 5;    Indels 0;    Gaps 0;  
Db    1040 PRGPAGPNGADGPQG 1054

RESULT 29

A44982  
collagen UC011 - pig roundworm (fragment)  
C;Species: Ascaris suum (pig roundworm)  
C;Date: 14-May-1993 #sequence\_revision 14-May-1993 #text\_change 20-Sep-1999  
C;Accession: AA4982  
R;Kingston, I.B.; Weinwright, S.M.; Cooper, D.  
Mol. Biochem. Parasitol. 37, 137-146, 1989  
A;Title: Comparison of collagen gene sequences in Ascaris suum and Caenorhabditis ele  
A;Reference number: A44982; MUID:90136710  
A;Accession: AA4982  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-228 <XIN>  
A;Cross-references: GB:J04657; NID:9159656; PIDN:AA29371.1; PID:9159657  
C;Superfamily: unassigned collagens

Query Match    58.7%;    Score 54;    DB 2;    Length 228;  
Best Local Similarity    66.7%;    Pred. No. 1.3;  
Matches 10;    Conservative 1;    Mismatches 4;    Indels 0;    Gaps 0;  
Db    41 PPGPPGPDKGDPG 55

RESULT 30

T20177  
hypothetical protein C53B4.5 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 15-oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 21-Jan-2000  
C;Accession: T20177  
R;Barks, M.  
submitted to the EMBL Data Library, December 1995  
A;Reference number: Z19233  
A;Accession: T20177  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-289 <NTL>  
A;Cross-references: EMBL:Z68215; PIDN:CAA92453.1; GSPDB:GN00022; CESP:C53B4.5  
A;Experimental source: clone C53B4  
C;Genetics:  
A;Gene: CESP:C53B4.5  
A;Map position: 4  
C;Superfamily: unassigned collagens

Query Match    58.7%;    Score 54;    DB 2;    Length 289;  
Best Local Similarity    66.7%;    Pred. No. 1.6;  
Matches 10;    Conservative 0;    Mismatches 5;    Indels 0;    Gaps 0;  
Db    178 PAGPPGPGKPGPAG 192

Search completed: November 1, 2002, 12:54:19

Fri Nov 1 13:09:33 2002

Job time : 15.5 secs

---

us-09-529-691a-3.rpr

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: November 1, 2002, 12:16:05 ; Search time 7 Seconds

Perfect score: US-09-529-691A-3  
Sequence: 1 PAGPMGPNGKDGVG 15

82.970 Million cell updates/sec

Title: US-09-529-691A-3

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SWISSPROT\_40;\*

Prd. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	66	71.7	1364	1 CA21_BOVIN	P02465 bos taurus
2	64	69.6	1356	1 CA21_ONCWM	09384 oncorhynchus
3	63	68.5	1366	1 CA21_CANFA	046392 canis familiaris
4	63	68.5	1366	1 CA21_HUMAN	03823 homo sapiens
5	62	67.4	526	1 CA21_RABIT	P08168 oryctolagus cuniculus
6	60	65.2	1372	1 CA21_MOUSE	Q01149 mus musculus
7	60	65.2	1372	1 CA21 RAT	P02466 rattus norvegicus
8	60	65.2	2944	1 CA17_HUMAN	P02388 homo sapiens
9	57	62.0	1459	1 CA12_MOUSE	P28481 mus musculus
10	56	60.9	369	1 CA12_CHICK	P02460 gallus gallus
11	56	60.9	636	1 CA13 RAT	P13341 rattus norvegicus
12	56	60.9	1804	1 CA1B_MOUSE	Q61245 mus musculus
13	56	60.9	1806	1 CA1B_HUMAN	P12107 homo sapiens
14	55	59.8	1262	1 CA13_CHICK	P1205 gallus gallus
15	55	59.8	1362	1 CA21_CHICK	P02467 gallus gallus
16	55	59.8	1418	1 CA12_HUMAN	P02458 homo sapiens
17	55	59.8	1453	1 CA11_CHICK	P02457 gallus gallus
18	55	59.8	1464	1 CA13_MOUSE	P08121 mus musculus
19	55	59.8	1650	1 CA2B_MOUSE	P064739 mus musculus
20	55	59.8	1736	1 CA2B_HUMAN	P11942 homo sapiens
21	54	58.7	305	1 YK61_CAEEL	P34340 caenorhabditis elegans
22	54	58.7	1049	1 CA13_BOVIN	P04258 bos taurus
23	54	58.7	1355	1 CA21_RANCA	O42250 ranas catesbeiana
24	54	58.7	1466	1 CA13_HUMAN	P02461 homo sapiens
25	54	58.7	1527	1 CA1H_MOUSE	P39061 mus musculus
26	54	58.7	1838	1 CA15_HUMAN	P20308 homo sapiens
27	53	57.6	307	1 CC36_CAEEL	P3403 caenorhabditis elegans
28	53	57.6	329	1 CC05_CAEEL	P18331 caenorhabditis elegans
29	53	57.6	374	1 PSPD_MOUSE	P5004 mus musculus
30	53	57.6	374	1 PSPD_RAT	P35248 rattus norvegicus
31	53	57.6	779	1 CAL1_BOVIN	P02453 bos taurus
32	53	57.6	1496	1 CA25_HUMAN	P03997 homo sapiens
33	56.5	301	1 CC02_CAEEL	P17556 caenorhabditis elegans	

## ALIGNMENTS

RESULT 1	CA21_BOVIN	ID CA21_BOVIN	STANDARD:	PRT; 1364 AA.
		ID	P02465	bos taurus
		AC	P02465;	062649;
		DT	21-JUL-1986	(Rel. 01, Created)
		DT	30-MAY-2000	(Rel. 39, Last sequence update)
		DT	16-OCT-2001	(Rel. 40, Last annotation update)
		DE	Collagen alpha 2(I) chain precursor.	COLLA2.
		GN		
		OC	Bovidae;	
		OC	Mammalia; Eutheria; Cetartiodactyla;	
		OC	Ruminantia; Pecora; Bovidae;	
		OX		
		RN	NCBI_TaxID=9913;	
		[1]		
		RP	SEQUENCE FROM N.A.	
		RC	MEDLINE=9829019; PubMed=9628255;	
		RX	Shimai T., Hattori S., Sakeuchi M., Inouye S., Kimura A., Ebihara T., Irie S., Nagai Y., Hori H.; "The complete cDNA coding sequence for the bovine proalpha2(I) chain of type I procollagen;" Matrix Biol. 17:85-88(1998).	
		RL		
		RN	"[2]"	
		RP	SEQUENCE OF 80-98.	
		RC	TISSUE=SKIN;	
		RX	MEDLINE=75036115; PubMed=4609475;	
		RA	Fieczek P.P., Breitkreitz D., Kuehn K.; "Amino acid sequence of the amino-terminal region of calf skin collagen;" Biochim. Biophys. Acta 365:305-310(1974).	
		RT		
		RL		
		RN	"[3]"	
		RP	SEQUENCE OF 95-415, AND REVISION.	
		RC	TISSUE=SKIN;	
		RX	MEDLINE=76091874; PubMed=173531;	
		RA	Fieczek P.P., Rexrodt F.W.; "The covalent structure of collagen. The amino-acid sequence of alpha2-CB4 from calf-skin collagen;" Eur. J. Biochem. 59:113-118(1975).	
		RT		
		RL		
		RN	"[4]"	
		RC	SEQUENCE OF 416-445.	
		RC	TISSUE=SKIN;	
		RC	MEDLINE=75008198; PubMed=4412529;	
		RA	Fieczek P.P., Furthmayr H., Kuehn K.; "Comparative sequence studies on alpha2-CB2 from calf, human, rabbit and pig-skin collagen;" Eur. J. Biochem. 47:257-261(1974).	
		RT		
		RL		
		RN	"[5]"	
		RC	SEQUENCE OF 446-481.	
		RC	TISSUE=SKIN;	
		RA	MEDLINE=75059350; PubMed=4435743;	
		RT	Fieczek P.P., Kuehn K.; "The covalent structure of collagen: amino acid sequence of the N-terminal region of alpha2-CB3 from rat skin collagen and alpha2-CB3 from calf skin collagen;" Hoppe-Seyler's Z. Physiol. Chem. 355:647-650(1974).	
		RL		

CC -!- FUNCTION: TYPE I COLLAGEN IS A MEMBER OF GROUP I COLLAGEN  
 CC (FIBRILLAR FORMING COLLAGEN).  
 CC SUBUNIT: TRIMERS OF ONE ALPHA 2(I) AND TWO ALPHA 1(I) CHAINS.  
 CC -!- TISSUE SPECIFICITY: FORMS THE FIBRILS OF TENDON, LIGAMENTS AND  
 CC BONES. IN BONES THE FIBRILS ARE MINERALIZED WITH CALCIUM  
 CC HYDROXYAPATITE.  
 CC -!- PFM: PROLINES AT THE THIRD POSITION OF THE TRIPPLET REPEATING  
 UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to license@isb-sib.ch).  
 CC  
 DR EMBL; AB006683; BAA25171\_1; -.  
 DR PIR: A02866;  
 DR InterPro; IPR00007; Collagen.  
 DR Pfam; PF01410; COLFI; 1.  
 DR Pfam; PF01391; Collagen; 17.  
 DR Prodrom; PDB02078; Fib\_collagen\_C; 1.  
 DR SMART; SM00038; COLFI; 1.  
 KW extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
 KW collagen; Signal.  
 FT SIGNAL 1 24 POTENTIAL.  
 FT PROPEP 25 79 AMINO-TERMINAL PROTEPTIDE.  
 FT CHAIN 80 1100 COLLAGEN ALPHA 2(I) CHAIN.  
 FT PROPEP 1101 1364 PYRROLIDONE CARBOXYLIC ACID.  
 FT MOD\_RES 80 80 CONVERTED TO AN ALDEHYDE GROUP THAT IS  
 INVOLVED IN CROSS-LINKING (PROBABLE).  
 FT MOD\_RES 84 84 HYDROXYLATION (PROBABLE).  
 FT MOD\_RES 175 175 HYDROXYLATION (PROBABLE).  
 FT MOD\_RES 196 196 HYDROXYLATION (PROBABLE).  
 FT MOD\_RES 262 262 HYDROXYLATION (PROBABLE).  
 FT MOD\_RES 307 307 HYDROXYLATION (PROBABLE).  
 FT MOD\_RES 352 352 HYDROXYLATION (PROBABLE).  
 FT CONFLICT 157 157 V -> P (IN REF. 3).  
 FT CONFLICT 187 187 K -> T (IN REF. 3).  
 FT CONFLICT 211 211 T -> K (IN REF. 3).  
 FT CONFLICT 298 300 PGA -> AGP (IN REF. 3).  
 FT CONFLICT 423 424 AT -> TA (IN REF. 4).  
 SQ SEQUENCE 1364 AA; 129064 MW; 5593F4D6B9ED19A CRC64;

Query Match Similarity 71.7%; Score 66; DB 1; Length 1364;  
 Best Local Similarity 73.3%; Pred. No. 0.13; Gaps 0;  
 Matches 11; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 PAGPNPGNGDKGVG 15  
 Db 1053 PAGPSGPAGKDGKIG 1067

---

RESULT 2  
 CA21\_ONCMY ID CA21\_ONCMY STANDARD; PRT; 1356 AA.  
 AC 093484;  
 DT 15-Oct-2001 (Rel. 40, Created)  
 DT 16-Oct-2001 (Rel. 40, Last sequence update)  
 DT 16-Oct-2001 (Rel. 40, Last annotation update)  
 DE collagen alpha 2(I) chain precursor.  
 GN COLIA2.  
 OS Oncorhynchus mykiss (Rainbow trout) (*Salmo gairdneri*).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Oncorhynchus.  
 OC NCBI\_TaxID=8022;  
 [1] SEQUENCE FROM N.A.

Qy 1 PAGPNPGNGDKGVG 15  
 Db 1053 PAGPSGPAGKDGKIG 1067

---

RESULT 3  
 CA21\_CANFA ID CA21\_CANFA STANDARD; PRT; 1366 AA.  
 AC 046392;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-May-2000 (Rel. 39, Last sequence update)  
 DT 16-Oct-2001 (Rel. 40, Last annotation update)  
 DE Collagen alpha 2(I) chain precursor.  
 GN COLIA2.  
 OS Canis familiaris (dog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
 OC NCBI\_TaxID=9615;  
 [1] SEQUENCE FROM N.A.

Qy 1 PAGPNPGNGDKGVG 15  
 Db 1049 PAGPHGPPGKDGRAG 1063

CC -!- FUNCTION: TYPE I COLLAGEN IS A MEMBER OF GROUP I COLLAGEN  
 CC (FIBRILLAR FORMING COLLAGEN).  
 CC -!- SUBUNIT: TRIMERS OF ONE ALPHA 2(I) AND TWO ALPHA 1(I) CHAINS.  
 CC [3]  
 CC -!- TISSUE SPECIFICITY: FORMS THE FIBRILS OF TENDON, LIGAMENTS AND  
 CC BONES. IN BONES THE FIBRILS ARE MINERALIZED WITH CALCIUM  
 CC HYDROXYAPATITE.  
 CC -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPIDE REPEATING  
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
 CC .....  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to license@lsb-sib.ch).  
 CC .....  
 CC EMBL; AF035120; AC:64485; 1;  
 DR InterPro; IPR00087; Collagen.  
 DR InterPro; IPR00085; Fib\_collagen\_C.  
 DR Pfam; PF01410; COLFI; 1.  
 DR Pfam; PF01391; Collagen; 18.  
 DR Prodom; PD002078; Fib\_collagen\_C; 1.  
 DR SMART; SM0038; COLFI; 1.  
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
 KW Collagen; Signal.  
 FT SIGNAL 1 24 POTENTIAL,  
 FT PROPEP 25 79 AMINO-TERMINAL PROPEPTIDE (BY  
 FT CHAIN 80 1102 COLLAGEN ALPHA 2(I) CHAIN.  
 FT PROPEP 1103 1366 CARBOXYL-TERMINAL PROPEPTIDE (BY  
 FT MOD\_RES 80 80 PYRROLIDONE CARBOXYLIC ACID (BY  
 FT MOD\_RES 84 84 SIMILARITY). CONVERTED TO AN ALDEHYDE GROUP THAT IS  
 FT INVOLVED IN CROSS-LINKING (BY SIMILARITY).  
 SQ SEQUENCE 1366 AA; 129400 MW; CD93696E080BCD4 CRC64;  
 Query Match 68 5%; Score 63; DB 1; Length 1366;  
 Beat Local Similarity 73.3%; Pred. No. 0.33; RT  
 Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0; RN  
 Db 1055 PAGPSGPAGKDGRTG 1069  
 RESULT 4  
 CAA1\_HUMAN STANDARD; PRT: 1366 AA.  
 ID CAA1\_HUMAN STANDAR  
 AC P08123; P02464; Q9UEB6; Q9UPH0;  
 DT 01-AUG-1988 (Rel. 08, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DE Collagen alpha 2(I) chain precursor.  
 GN COL1A2.  
 OS Homo sapiens (Human).  
 OC Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homindae; Homo.  
 OX NCBI\_TAXID=9606;  
 RN [1] SEQUENCE FROM N.A.  
 RP MEDLINE=8058462; PubMed=2824475;  
 RN de Wet W.J., Bernard M.P., Benson-Chanda V., Chu M.-L., Dickson L.A.,  
 RA Weil D., Ramirez F.,  
 RT "Organization of the human pro-alpha 2(I) collagen gene.",  
 RL J. Biol. Chem. 262:16032-16036(1987).  
 RN [2] SEQUENCE FROM N.A.  
 RP Korkko J.M., Earley J.J., Ala-Kokko L., Prockop D.J.;  
 RA "Analysis of the COL1 and COL1A2 genes by CSGE and DNA sequencing in  
 RT 14 patients with mild or (Type I). Identification of common sequences  
 RT for null allele mutations.",  
 RT Submitted (MAY-1997) to the EMBL/GenBank/DDJB databases.  
 RN SEQUENCE OF 1-765 FROM N.A.  
 RC TISSUE=Placenta;  
 RX MEDLINE=8339824; PubMed=3421913;  
 RA Kulvanieni H., Tromp G., Chu M.-L., Prockop D.J.;  
 RT "Structure of a full-length cDNA clone for the prepro alpha 2(I)  
 chain of human type I procollagen. Comparison with the chicken gene  
 RT confirms unusual patterns of gene conservation.";  
 RL Biochem. J. 252:633-640(1988).  
 RN [4] SEQUENCE OF 181-1366 FROM N.A.  
 RA Kalicki J., Wamsley P., Gibson A.;  
 RT Submitted (SEP-1997) to the EMBL/GenBank/DDJB databases.  
 RN [5] SEQUENCE OF 623-1366 FROM N.A.  
 RX MEDLINE=8317819; PubMed=6687691;  
 RA Bernard M.P., Myers J.C., Chu M.-L., Ramirez F., Eikenberry E.F.,  
 RT Prockop D.J.;  
 RA Click E.M., Bernstein P.;  
 RT "Isolation and characterization of the cyanoen bromide peptides from  
 RT procollagen. Comparison with chick cDNA for pro alpha 2(I) identifies  
 RT structurally conserved features of the protein and the gene.";  
 RL Biochemistry 22:1139-1145(1983).  
 RN [6] SEQUENCE OF 80-96.  
 RP TISSUE=Skin;  
 RX MEDLINE=7103825; PubMed=5529814;  
 RA Click E.M., Bernstein P.;  
 RT "Isolation and characterization of the cyanoen bromide peptides from  
 RT the alpha 1 and alpha 2 chains of human skin collagen.";  
 RL Biochemistry 9:4699-4706(1970).  
 RN [7] SEQUENCE OF 417-447.  
 RP TISSUE=Skin;  
 RC MEDLINE=75008198; PubMed=4412529;  
 RX Fietzek P.P., Furthmayr H., Kuehn K.;  
 RT "Comparative sequence studies on alpha2-CB2 from calf, human, rabbit  
 RT and pig-skin collagen.";  
 RL Eur. J. Biochem. 47:257-261(1974).  
 RN [8] SEQUENCE OF 145-198 FROM N.A.  
 RP MEDLINE=82498792; PubMed=3403536;  
 RA Kulvanieni H., Sabol C., Tromp G., Sippola-Thiele M., Prockop D.J.;  
 RT "A 19-base pair deletion in the pro-alpha 2(I) gene of type I  
 RT procollagen that causes in-frame RNA splicing from exon 10 to exon 12  
 RT in a proband with atypical osteogenesis imperfecta and in his  
 RT asymptomatic mother.";  
 RT J. Biol. Chem. 263:11407-11413(1988).  
 RL [9] SEQUENCE OF 960-1351 FROM N.A.  
 RN TISSUE=Skin;  
 RX MEDLINE=90304220; PubMed=2364107;  
 RA Maekiae J.K., Vuorio T., Vuorio E.;  
 RT "Growth-dependent modulation of type I collagen production and mRNA  
 RT levels in cultured human skin fibroblasts.";  
 RL Biochim. Biophys. Acta 1049:171-176(1990).  
 RN [10] REVIEW ON VARIANTS.  
 RP MEDLINE=91184577; PubMed=2010058;  
 RA Kulvanieni H., Tromp G., Prockop D.J.;  
 RT "Mutations in collagen genes: causes of rare and some common diseases  
 RT in humans";  
 RL FASEB J. 5:2052-2060(1991).  
 RN [11] REVIEW ON VARIANTS.  
 RX MEDLINE=9725559; PubMed=9101290;  
 RA Kulvanieni H., Tromp G., Prockop D.J.;  
 RT "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-  
 associated collagen (type IX), and network-forming collagen (type X)  
 RT cause a spectrum of diseases of bone, cartilage, and blood vessels.";  
 RL Hum. Mutat. 9:300-315(1997).  
 RN [12]

RP	REVIEW ON OI VARIANTS.	RX
RX	MEDLINE=51374476; PubMed=1895312;	RN
RA	Byers P.H., Wallis G.A., Willing M.C.;	RN
RT	"Osteogenesis imperfecta: translation of mutation to phenotype.";	RN
J. Med. Genet.	28:433-442(1991).	[13]
RL		
RN	REVIEW ON OI VARIANTS.	RX
RX	MEDLINE=97159389; PubMed=2016532;	RN
RA	Dalglish R.;	RN
RT	"The human type I collagen mutation database.";	RN
RL	Nucleic Acids Res.	[14]
RN		
RP	VARIANT EDS-VII-A2.	RX
RX	MEDLINE=8805913; PubMed=3680255;	RN
RA	Wirtz M.K., Gianville R.W., Steinmann B., Rao V.H., Hollister D.W.;	RN
RT	"Ehlers-Danlos syndrome type VII B. Deletion of 18 amino acids comprising the N-telopeptide region of a pro-alpha 2(I) chain.";	RN
RT	J. Biol. Chem.	[15]
RL	262:16376-16385(1987).	
RN		
RP	SEQUENCE OF 1990-1107 FROM N.A. AND VARIANT OI-IV ARG-1102.	RX
RX	MEDLINE=88227975; PubMed=2897303;	RN
RA	Wenstrup R.J., Cohn D.H., Cohen T., Byers P.H.;	RN
RT	"Arginine for glycine substitution in the triple-helical domain of the products of one alpha 2(I) collagen allele (COLIA2) produces the osteogenesis imperfecta type IV phenotype.";	RN
RT	J. Biol. Chem.	[16]
RL	263:7734-7740(1988).	
RN		
RP	VARIANT OI-II ASP-997.	RX
RX	MEDLINE=89123407; PubMed=2914942;	RN
RA	Baldwin C.T., Constantino C., Dumars K.W., Prockop D.J.;	RN
RT	"A single base mutation that converts glycine 907 of the alpha 2(I) chain of type I procollagen to aspartate in a lethal variant of osteogenesis imperfecta. The single amino acid substitution near the carboxyl terminus destabilizes the whole triple helix.";	RN
RT	J. Biol. Chem.	[17]
RL	264:3002-3006(1989).	
RN		
RP	VARIANT OI-II SER-955.	RX
RX	MEDLINE=893805; PubMed=277764;	RN
RA	Lamande S.R., Dahl H.-H.M., Cole W.G., Bateman J.F.;	RN
RT	"Characterization of point mutations in the collagen COLIA1 and COLIA2 genes causing lethal perinatal osteogenesis imperfecta.";	RN
RT	J. Biol. Chem.	[18]
RL	264:15809-15812(1989).	
RN		
RP	VARIANT OI-II CYS-199.	RX
RX	MEDLINE=90368825; PubMed=2394758;	RN
RA	Fertala A., Westerhausen A., Morris G.M., Rooney J.E., Prockop D.J.;	RN
RT	"Two cysteine substitutions in the type I procollagen genes (COLIA1 and COLIA2) that cause lethal osteogenesis imperfecta. The location of glycine substitutions does not in any simple way predict their effects on protein function or phenotype.";	RN
RT	J. Hum. Genet.	[19]
RL	47:A216-A216(1990).	
RN		
RP	VARIANT EDS-VII-A2.	RX
RX	MEDLINE=90368825; PubMed=2394758;	RN
RA	Weil D., D'Alessio M., Ramirez F., Eyre D.R.;	RN
RT	"Structural and functional characterization of a splicing mutation in the pro-alpha 2(I) collagen gene of an Ehlers-Danlos type VII patient.";	RN
RT	J. Biol. Chem.	[20]
RL	265:16007-16011(1990).	
RN		
RP	VARIANT OI-IV VAL-676.	RX
RX	MEDLINE=91291136; PubMed=2046412;	RN
RA	Bateman J.F., Hannigan M., Chan D., Cole W.G.;	RN
RT	"Characterization of a type I collagen alpha 2(I) glycine-586 to valine substitution in osteogenesis imperfecta type IV. Detection of the mutation and prenatal diagnosis by a chemical cleavage method.";	RN
RT	Biochem. J.	[21]
RL	276:765-770(1991).	
RN		
RP	VARIANTS OI CYS-349 AND CYS-736.	RX
RX	MEDLINE=911589; PubMed=199009;	RN
RA	Wenstrup R.J., Shrago-Howe A.W., Phillips C.L.,	RN
RT	Byers P.H., Cohn D.H.;	RN
RT	"The effects of different cysteine for glycine substitutions within alpha 2(I) chains. Evidence of distinct structural domains within the	
RT		
RT	type I collagen triple helix.";	RN
RT	J. Biol. Chem.	[22]
RT	266:2590-2594(1991).	
RP	VARIANT OI-II ARG-784.	RX
RX	MEDLINE=9140689; PubMed=1874719;	RN
RA	Tsuneyoshi T., Westerhausen A., Constantinou C.D., Prockop D.J.;	RN
RT	"Substitutions for glycine alpha 1-637 and glycine alpha 2-694 of type I procollagen in lethal osteogenesis imperfecta. The conformational strain on the triple helix introduced by a glycine substitution can be transmitted along the helix.";	RN
RT	J. Biol. Chem.	[23]
RL	266:15608-15613(1991).	
RN		
RP	VARIANT OI-IV SER-751.	RX
RX	MEDLINE=91271401; PubMed=2052622;	RN
RA	Spotila L.D., Constantinou C.D., Sereda L., Ganguly A., Riggs B.L., Prockop D.J.;	RN
RA	"Mutation in a gene for type I procollagen (COLIA2) in a woman with postmenopausal osteoporosis: evidence for phenotypic and genotypic overlap with mild osteogenesis imperfecta.";	RN
RT	Proc. Natl. Acad. Sci. U.S.A. 88:5423-5427(1991).	RN
RT	[24]	
RP	VARIANT OI-II ARG-547.	RX
RX	MEDLINE=9344832; PubMed=1284475;	RN
RA	Bateman J.F., Moeller I., Hannigan M., Chan D., Cole W.G.;	RN
RT	"Lethal perinatal osteogenesis imperfecta due to a type I collagen alpha 2(I) GLY to ARG substitution detected by chemical cleavage of mRNA:cDNA sequence mismatch.";	RN
RT	Hum. Mutat. 1:55-62(1992).	RN
RP	VARIANT OI-II ASP-670.	RX
RX	MEDLINE=93054637; PubMed=1385413;	RN
RT	Query Match Best Local Similarity 68.5%; Score 63; DB 1; Length 1366; Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;	RN
OY	1 PAGPWPKNGKQKG 15	RN
Db	1055 PAGPSOPAGKDGRG 1069	RN
RESULT 5		
CA21 RABIT		
ID CA21 RABIT	STANDARD;	PRT;
AC 028668;		526 AA.
DT 16-CT-2001 (Rel. 40. Created)		
DT 16-OCT-2001 (Rel. 40. Last sequence update)		
DT 16-OCT-2001 (Rel. 40. Last annotation update)		
DE Collagen alpha 2(I) chain precursor (Fragment).		
GN COLIA2.		
OS Oryctolagus cuniculus (Rabbit).		
OC Dukarwota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Butheria; Diagonomorpha; Leporidae; Oryctolagus.		
OC NCBI_TaxID:9986;		
RN [1]		
RP SEQUENCE FROM N.A.		
RT STRAIN="New Zealand white; TISSUE="Calvaria;		
RT Inoue S., Okazaki T.;		
RT "Alpha 2 type I collagen gene expression in the rabbit knee ligaments: variations during the newborn development and in the adult age.";		
RL Biomed. Res. 16:219-227(1995).		
CC -1- FUNCTION: TYPE I COLLAGEN IS A MEMBER OF GROUP I COLLAGEN		
CC (FIBRILLAR FORMING COLLAGEN)		
CC -1- SUBUNIT: TRIMERS OF ONE ALPHA 2(I) AND TWO ALPHA 1(I) CHAINS.		
CC -1- TISSUE SPECIFICITY: FORMS THE FIBRILS OF TENDON, LIGAMENTS AND BONES. IN BONES THE FIBRILS ARE MINERALIZED WITH CALCIUM HYDROXYAPATITE.		
CC -1- PROLINES AT THE THIRD POSITION OF THE TRIPETIDE REPEATING UNIT (G-X-Y) ARE HYDROXILATED IN SOME OR ALL OF THE CHAINS.		
CC		
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation of the European Bioinformatics Institute. There are no restrictions on its		

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: De9399; BAAB8391; -  
DR InterPro; IPR00007; Collagen.  
DR InterPro; IPR00085; Fib.collagen\_C.  
DR Pfam; PF01410; COLFI; 1.  
DR Prodom; PD002078; Fib.collagen\_C; 1.  
DR SMART; SW00038; COLFI; 1.  
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
KW Collagen.  
FT CHAIN <1 1 262 COLLAGEN ALPHA 2(I) CHAIN.  
FT PROPEP 263 526 CARBOXYL-TERMINAL PROPEPTIDE  
SQ SEQUENCE 526 AA; 53129 MW; (BY SIMILARITY); 7CA8FORA6953846D CRC64;  
Query Match 67.4%; Score 62; DB 1; Length 526;  
Best Local Similarity 73.3%; Pred. No. 0.19;  
Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
OY 1 PAGPWGPNGKDGKVG 15 DR SMART; SW00038; COLFI; 1.  
Db 215 PAGSGPVGKDGRS 229 DR Pfam; PF01410; COLFI; 1.  
DR InterPro; IPR00085; Fib.collagen\_C.  
DR Prodom; PD002078; Fib.collagen\_C.  
DR SMART; SW00038; COLFI; 1.  
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
KW Glycoprotein; Collagen; Signal;  
FT SIGNAL 1 22 POTENTIAL.  
FT PROPEP 23 85 AMINO-TERMINAL PROPEPTIDE  
FT CHAIN 86 1108 (BY SIMILARITY);  
FT PROPEP 1109 1372 COLLAGEN ALPHA 2(I) CHAIN.  
FT MOD\_RES 86 86 CARBOXYL-TERMINAL PROPEPTIDE  
FT FT CONFLICT 1273 1273 PYRROLIDINE CARBOXYLIC ACID (BY  
FT MOD\_RES 90 90 CONFLICT) SIMILARITY).  
FT FT CONFLICT 1167 1167 R -> TT (IN REF. 1);  
FT SEQUENCE 1372 AA; 129557 MW; (BY SIMILARITY);  
Query Match 65.2%; Score 60; DB 1; Length 1372;  
Best Local Similarity 73.3%; Pred. No. 0.84;  
Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
OY 1 PAGPWGPNGKDGKVG 15 DR N-LINKED (GLCNAC. . .) (POTENTIAL).  
Db 1061 PAGSGPVGKDGRS 1075 V -> A (IN REF. 4);  
SEQUENCE FROM N.A.  
RC TISSUE=Salvaria;  
RX MEDLINE=92372043; Pubmed=1505972;  
RA Phillips C.L., Morgan A.L., Lever L.W., Wenstrup R.J.;  
RT "Sequence analysis of a full-length cDNA for the murine pro alpha  
2(I) collagen chain: comparison of the derived primary structure with  
human pro alpha 2(I) collagen.";  
RL Genomics 13:1345-1346(1992).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=breast tumor;  
RA Strausberg R.;  
RL Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.  
RN [3]  
RP SEQUENCE OF 1-110 FROM N.A.  
RC TISSUE=Salvaria;  
RX MEDLINE=92084969; Pubmed=1748823;  
RA Phillips C.L., Lever L.W., Pinnell S.R., Quarles L.D.,  
RA Wenstrup R.J.;  
RT "Construction of a full-length murine pro alpha 2(I) collagen cDNA by  
the polymerase chain reaction.";  
RL J. Invest. Dermatol. 97:980-984(1991).  
RN [4]  
RP SEQUENCE OF 1-23 FROM N.A.  
RX MEDLINE=97289650; Pubmed=3039494;  
RA Rossi P., de Crombrugge B.;  
RT "Identification of a cell-specific transcriptional enhancer in the  
first intron of the mouse alpha 2 (type I) collagen gene.";  
RL Proc. Natl. Acad. Sci. U.S.A. 84:5905-594(1987).  
RN [5]  
RP SEQUENCE OF 86-98.

CC (FIBRILLAR FORMING COLLAGEN).  
CC -- SUBUNIT: TRIMERS OF ONE ALPHA 2(I) AND TWO ALPHA 1(I) CHAINS.  
CC -- TISSUE SPECIFICITY: FORMS THE FIBRILS OF TENDON, LIGAMENTS AND  
CC BONES. IN BONES THE FIBRILS ARE MINERALIZED WITH CALCIUM  
CC HYDROXYAPATITE.  
CC -- PTM: PROLINES AT THE THIRD POSITION OF THE TRipeptide REPEATING  
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to license@isb-sib.ch).  
CC -----  
CC EMBL: X88251; CAA41205\_1; -  
DR EMBL; K01832; AAA37331; -.  
DR PIR; A43291; A43291.  
DR MGI; MGI-88468; Colla2.  
DR InterPro; IPR00087; Collagen.  
DR Prodom; PD002078; Fib.collagen\_C.  
DR SMART; SW00038; COLFI; 1.  
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
KW Glycoprotein; Collagen; Signal;  
FT SIGNAL 1 22 POTENTIAL.  
FT PROPEP 23 85 AMINO-TERMINAL PROPEPTIDE  
FT CHAIN 86 1108 (BY SIMILARITY);  
FT PROPEP 1109 1372 COLLAGEN ALPHA 2(I) CHAIN.  
FT MOD\_RES 86 86 CARBOXYL-TERMINAL PROPEPTIDE  
FT FT CONFLICT 1273 1273 PYRROLIDINE CARBOXYLIC ACID (BY  
FT MOD\_RES 90 90 CONFLICT) SIMILARITY).  
FT FT CONFLICT 1167 1167 R -> TT (IN REF. 1);  
FT SEQUENCE 1372 AA; 129557 MW; (BY SIMILARITY);  
Query Match 65.2%; Score 60; DB 1; Length 1372;  
Best Local Similarity 73.3%; Pred. No. 0.84;  
Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
OY 1 PAGPWGPNGKDGKVG 15 DR N-LINKED (GLCNAC. . .) (POTENTIAL).  
Db 1061 PAGSGPVGKDGRS 1075 V -> A (IN REF. 4);  
SEQUENCE FROM N.A.  
RC TISSUE=RAT;  
RX MEDLINE=92466; OPR158; STANDARD;  
RA P02466; OPR158; STANDARD;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Collagen alpha 2(I) chain precursor.  
GN COLIA2.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrates; Buteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciuromorphi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Guenther D., Seibold S., Marx M.;  
RL Submitted (JAN-1999) to the EMBL/GenBank/DDBJ databases.  
RN [2]  
RP SEQUENCE OF 86-98.

RC TISSUE=Skin;  
 RX MEDLINE=6712628; PubMed=5337885; Fib-collagen\_C.  
 RA Kang A.H.; Bornstein P.; Piez K.A.;  
 RT "The amino acid sequence of peptides from the cross-linking region of  
 rat skin collagen.";  
 RL Biochemistry 67:88-195 (1967).  
 [3]  
 RP SEQUENCE OF 99-102.  
 RC TISSUE=Skin;  
 RX MEDLINE=69206881; PubMed=5785232;  
 RA Fietzek P.P.; Piez K.A.;  
 RT "Isolation and characterization of the cyanogen bromide peptides from  
 the alpha 2 chain of rat skin collagen.";  
 RL Biochemistry 8:2129-2133(1969).  
 [4]  
 RP SEQUENCE OF 102-144.  
 RC TISSUE=Skin;  
 RX MEDLINE=73049496; PubMed=4636752;  
 RA Fietzek P.P.; Kell I.; Kuehn K.;  
 RT "The covalent structure of collagen. Amino acid sequence of the N-  
 terminal region of alpha 2-CB4 from calf and rat skin collagen.";  
 RL FEBS Lett. 26:66-68(1972).  
 [5]  
 RP SEQUENCE OF 423-452.  
 RC TISSUE=Skin;  
 RX MEDLINE=71115216; PubMed=5544653;  
 RA Highberger J.H.; Kang A.H.; Gross J.;  
 RT "Comparative studies on the amino acid sequence of the alpha 2-CB2  
 peptides from chick and rat skin collagens.";  
 RL Biochemistry 10:610-616(1971).  
 [6]  
 RP SEQUENCE OF 453-501.  
 RC TISSUE=Skin;  
 RX MEDLINE=75059250; PubMed=4435743;  
 RA Fietzek P.P.; Kuehn K.;  
 RT "The covalent structure of collagen: amino acid sequence of the N-  
 terminal region of alpha2-CB3 from rat skin collagen and alpha2-CB3.5  
 from calf skin collagen.";  
 RL Hoppe-Seyler's Z. Physiol. Chem. 355:647-650(1974).  
 [7]  
 RP SEQUENCE OF 791-836.  
 RC TISSUE=Skin;  
 RX MEDLINE=74055004; PubMed=4763308;  
 RA Fietzek P.P.; Kuehn K.;  
 RT "The covalent structure of collagen: amino acid sequence of the N-  
 terminal region of alpha 2-CB5 from rat skin collagen.";  
 RL FEBS Lett. 36:289-291(1973).  
 [8]  
 RP ORDER OF CNBR PEPTIDES.  
 RX MEDLINE=70181822; PubMed=5443712;  
 RA Vuust J.; Lane J.M.; Fietzek P.P.; Miller E.J.; Piez K.A.;  
 RT "The order of the CNBr peptides from the alpha 2 chain of collagen.";  
 RL Biochem. Biophys. Res. Commun. 38:703-708(1970).  
 CC -!- FUNCTION: TYPE I COLLAGEN IS A MEMBER OF GROUP I COLLAGEN  
 (FIBRILLAR FORMING COLLAGEN).  
 CC -!- SUBUNIT: TIMERS OF ONE ALPHA 2(I) AND TWO ALPHA 1(I) CHAINS.  
 CC -!- TISSUE SPECIFICITY: FORMS THE FIBRILS OF TENDON, LIGAMENTS AND  
 BONES. IN BONES THE FIBRILS ARE MINERALIZED WITH CALCIUM  
 HYDROXYAPATITE.  
 CC -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPETIDE REPEATING  
 UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.

.....

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR IntePro; IPR00885; Fib-collagen\_C.  
 DR Pfam; PF01391; Collagen; 18.  
 DR PRODOM; PD002078; Fib\_collagen\_C; 1.  
 DR SMART; SM00038; COLEI; 1.  
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
 KW Glycoprotein; Collagen; Signal.  
 FT SIGNAL 1 24 POTENTIAL.  
 FT PROPEP 25 AMINO-TERMINAL PROPEPTIDE  
 FT CHAIN 85 (BY SIMILARITY; COLLAGEN ALPHA 2(I) CHAIN.  
 FT PROPEP 1109 CARBOXY-TERMINAL PROPEPTIDE  
 FT MOD\_RES 86 CONVERTED TO AN ALDEHYDE GROUP THAT IS  
 FT MOD\_RES 90 INVOLVED IN CROSS-LINKING.  
 FT CARBOYD 1273 N-LINED (GLCNAC. . .) (POTENTIAL).  
 FT SITE 828 132 T -> P (IN REF. 4).  
 FT SITE 1011 132 S -> P (IN REF. 4).  
 FT SITE 1013 137 CELL ATTACHMENT SITE (POTENTIAL).  
 FT MOD\_RES 86 CELL ATTACHMENT SITE (POTENTIAL).  
 FT MOD\_RES 90 PYRROLIDONE CARBOXYLIC ACID (PROBABLE).  
 FT CARBOYD 1273 132 ST -> TS (IN REF. 5).  
 FT CONFFLICT 132 E -> Z (IN REF. 6).  
 FT CONFFLICT 137 N -> A (IN REF. 6).  
 FT CONFFLICT 145 MISSING (IN REF. 4).  
 FT CONFFLICT 432 R -> K (IN REF. 7).  
 FT CONFFLICT 494 497  
 FT CONFFLICT 502 790  
 FT CONFFLICT 825 825  
 SO SEQUENCE 1372 AA; 129564 MW; B069371A8DB20A72 CRC64;

Query Match Best Local Similarity 65.3%; Score 60; DB 1; Length 1372;  
 Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 PAGPWNPNKGKV 15  
 Db 1061 PAGPSQPIGDGRSG 1075

RESULT 8

CA17\_HUMAN ID CA17\_HUMAN  
 AC 002386; Q14054; Q16507; PRT; 2944 AA.  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-FEB-1995 (Rel. 33, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Collagen alpha 1(VII) chain precursor (Long-chain collagen) (LC  
 DE collagen).  
 GN COL7A1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarhini; Hominidae; Homo.  
 OX NCBI\_TAXID=9606;

RP SEQUENCE FROM N. A.

RX MEDLINE=94327508; PubMed=8051117;  
 RA Christiano A.M.; Greenspan D.S.; Lee S.; Utton J.;  
 RT "Cloning of human type VII collagen. Complete primary sequence of the  
 alpha 1(VII) chain and identification of intragenic polymorphisms.";  
 RL J. Biol. Chem. 269:20256-20262(1994).  
 RN [2]  
 RP SEQUENCE OF 128-1493 FROM N.A., AND PARTIAL SEQUENCE.  
 RX MEDLINE=9338417; PubMed=1307247;  
 RA Christiano A.M.; Rosenbaum L.M.; Chung-Honet L.C.; Parente M.G.,  
 RA Woodley D.T.; Pan T.C.; Zhang R.Z.; Chu M.-L.; Burgesson R.E.,  
 RA Utton J.;  
 RT "The large non-collagenous domain (NC-1) of type VII collagen is  
 amino-terminal and chimeric. Homology to cartilage matrix protein,  
 RT the type III domains of fibronectin and the A domains of von  
 RL Willebrand factor.";  
 RL Hum. Mol. Genet. 1:475-481(1992).  
 RN [3]  
 RP SEQUENCE OF 815-1439 FROM N.A.  
 RX MEDLINE=91334380; PubMed=1871109;

- RA Parente M.G., Chung L.C., Ryynänen J., Woodley D.T., Wynn K.W., Bauer E.A., Mattei M.-G., Chu M.-L., Utton J.; "Human type VII collagen: cDNA cloning and chromosomal mapping of the gene.;" Proc. Natl. Acad. Sci. U.S.A. 88:6931-6935(1991). [4]
- RN SEQUENCE OF 369-1255 FROM N.A.
- RX MEDLINE=93107742; PubMed=169284;
- RA Gaumon W.R., Abernethy M.L., Padilla K.M., Prisayanh P.S., Cook M.E., Wright J., Brigham R.A., Hunt S.W. III; "Noncollagenous (NCI) domain of collagen VII resembles multidomain adhesion proteins involved in tissue-specific organization of extracellular matrix."; J. Invest. Dermatol. 99:691-696(1992). [5]
- RN SEQUENCE OF 340-675 FROM N.A.
- RC TISSUE="keratinocytes";
- RX MEDLINE=92231902; PubMed=1567409;
- RA Tanaka T., Takahashi K., Furukawa F., Iimura S.; "Molecular cloning and characterization of type VII collagen cDNA.;" Biochem. Biophys. Res. Commun. 183:958-963(1992). [6]
- RN SEQUENCE OF 2395-2944 FROM N.A.
- RX MEDLINE=93271985; PubMed=8199916;
- RA Greenspan D.S.; "The carboxyl-terminal half of type VII collagen, including the non-collagenous NC-2 domain and intron/exon organization of the corresponding region of the COL7A1 gene.;" Hum. Mol. Genet. 2:273-278(1993). [7]
- RN SEQUENCE OF 1-87 FROM N.A.
- RC TISSUE="placenta";
- RX MEDLINE=94375010; PubMed=8088784;
- RA Christiano A.M., Hoffman G.G., Chung-Honet L.C., Lee S., Cheng W., Utton J., Greenspan D.S.; "Structural organization of the human type VII collagen gene (COL7A1), composed of more exons than any previously characterized gene.;" Genomics 21:169-179(1994). [8]
- RN REVIEW ON DEB VARIANTS.
- RX MEDLINE=98041696; PubMed=9375848;
- RA Jaakkola A., Puikkonen L., Utton J.; "Molecular basis of dystrophic epidermolysis bullosa: mutations in the type VII collagen gene (COL7A1).;" Hum. Mutat. 10:338-347(1997). [9]
- RN VARIANT RDEB LYS-2798.
- RX MEDLINE=93291877; PubMed=8513326;
- RA Christiano A.M., Greenspan D.S., Hoffman G.G., Zhang X., Tamai Y., Lin A.N., Dietz H.C., Hovnanian A., Utton J.; "A missense mutation in type VII collagen in two affected siblings with recessive dystrophic epidermolysis bullosa.;" Nat. Genet. 4:62-66(1993). [10]
- RN VARIANT DDEB SER-2040.
- RX MEDLINE=94224777; PubMed=8170945;
- RA Christiano A.M., Ryynänen M., Utton J.; "Dominant dystrophic epidermolysis bullosa: identification of a Gly->Ser substitution in the triple helical domain of type VII collagen.;" Proc. Natl. Acad. Sci. U.S.A. 91:3549-3553(1994). [11]
- RN VARIANT PEB-DDEB CYS-2623.
- RX MEDLINE=96081220; PubMed=8541842;
- RA Christiano A.M., Lee J.Y.-Y., Chen W.J., Laforgia S., Utton J.; "Pretilbial epidermolysis bullosa: genetic linkage to COL7A1 and identification of a glycine-to-cysteine substitution in the triple-helical domain of type VII collagen.;" Hum. Mol. Genet. 4:1579-1583(1995). [12]
- RN VARIANT DDEB ARG-2043.
- RX MEDLINE=95164985; PubMed=7861014;
- RA Christiano A.M., Morricone A., Paradisi M., Angelo C., Mazzanti C., Cavaliere R., Utton J.; "A glycine-toarginine substitution in the triple-helical domain of type VII collagen in a family with dominant dystrophic epidermolysis bullosa.;" J. Invest. Dermatol. 104:438-440(1995). [13]
- RN VARIANTS RDEB AND DDEB.
- RX MEDLINE=96220518; PubMed=8644729;
- RA Christiano A.M., McGrath J.A., Tan K.C., Utton J.; "Glycine substitutions in the triple-helical region of type VII collagen result in a spectrum of dystrophic epidermolysis bullosa phenotypes and patterns of inheritance.;" Am. J. Hum. Genet. 58:671-681(1996). [14]
- RN VARIANT RDEB ARG-2575.
- RX MEDLINE=9615068; PubMed=8592061;
- RA Shinilzu H., McGrath J.A., Christiano A.M., Nishikawa T., Utton J.; "Molecular basis of recessive dystrophic epidermolysis bullosa: genotype/phenotype correlation in a case of moderate clinical severity.;" J. Invest. Dermatol. 106:119-124(1996). [15]
- RN VARIANT RDEB ARG-1782.
- RX MEDLINE=96183562; PubMed=8618018;
- RA Christiano A.M., McGrath J.A., Utton J.; "Influence of the second COL7A1 mutation in determining the phenotypic severity of recessive dystrophic epidermolysis bullosa.;" J. Invest. Dermatol. 106:766-770(1996). [16]
- RN VARIANT RDEB ASP-2073.
- RX MEDLINE=96310789; PubMed=8757758;
- RA Dunnill M.G.S., McGrath J.A., Richards A.J., Christiano A.M., Utton J., Pope F.M., Eady R.A.J.; "Clinicopathological correlations of compound heterozygous COL7A1 mutations in recessive dystrophic epidermolysis bullosa.;" J. Invest. Dermatol. 107:171-177(1996). [17]
- RN VARIANTS RDEB W-1982; G-2008; A-2025; B-2049; G-2063 AND R-2075.
- RX MEDLINE=97465605; PubMed=9336325;
- RA Hovnanian A., Rochat A., Bodemer C., Petit B., Rivers C.A., Prost C., Fraitag S., Christiano A.M., Utton J., Lathrop M., Barrandon Y., de Post Y.; "Characterization of 18 new mutations in COL7A1 in recessive dystrophic epidermolysis bullosa provides evidence for distinct molecular mechanisms underlying defective anchoring fibril formation.;" Am. J. Hum. Genet. 61:599-610(1997). [18]
- RN VARIANT RDEB ARG-1652.
- RX MEDLINE=98106792; PubMed=9444387;
- RA Csernai-Mitelman P.B., Karpati S., Horvath A., Christiano A.M.; "Identification of a glycine substitution and a splice site mutation in the type VII collagen gene in a proband with mitis recessive dystrophic epidermolysis bullosa.;" Arch. Dermatol. Res. 289:640-645(1997). [19]
- RN VARIANTS DEB ARG-2009 AND ARG-2043.
- RX MEDLINE=9735688; PubMed=9215684;
- RA Wibberg J.-O., Hammami-Haussali N., Nilissen O., Anton-Lamprecht T., Navjor S.L., Kerbacher K., Zimmermann M., Krafcik P., Gedde-Dahl T. Jr., Bruckner-Tuderman L.; "Modulation of disease severity of dystrophic epidermolysis bullosa by a splice site mutation in combination with a missense mutation in the COL7A1 gene.;" Hum. Mol. Genet. 6:1125-1135(1997). [20]
- RN VARIANTS DDEB ASP-1519; ASP-2006; GLU-2015 AND ARG-2034.
- RX MEDLINE=98334662; PubMed=968111;
- RA Hammami-Haussali N., Schumann H., Raghunath M., Kilgus O., Luethi U., Luger T., Bruckner-Tuderman L.; "Some, but not all, glycine substitution mutations in COL7A1 result in intracellular accumulation of collagen VII, loss of anchoring

RT fibrils' and skin blistering";  
 RL J. Biol. Chem. 273:19228-19234(1998).  
 DR [21]  
 RP VARIANT DEB CYS-2008; ARG-2207 AND SER-2775.  
 RX MEDLINE=9410969; PubMed=9740253;  
 RA Kon A., Puukkinen L., Ishida-Yamamoto A., Hashimoto I., Ututto J.;  
 RT "Novel COL1A1 mutations in dystrophic forms of epidermolysis  
 bullousa";  
 RL J. Invest. Dermatol. 111:534-537(1998).  
 RN [22]  
 VARIANT RHEB ARG-1347.  
 RX MEDLINE=93019477; PubMed=9804332;  
 RA Terraccina M., Postoraro P., Schubert M., Sonego G., Atzori F.,  
 RA Zambruno G., Bruckner-Tuderman L., Castiglione D.;  
 RT Compound heterozygosity for a recessive glycine substitution and a  
 splice site mutation in the COL1A1 gene causes an unusually mild form  
 RT of localized recessive dystrophic epidermolysis bullosa.";  
 RL J. Invest. Dermatol. 111:744-750(1998).  
 RN [23]  
 RP VARIANT DEB TRP-2034; VAL-2040; ARG-2043; ARG-2064 AND ASP-2713.  
 RX MEDLINE=99072663; PubMed=9856843;  
 Query Match 65.2%; Score 60; DB 1; Length 2944;  
 Best Local Similarity 66.7%; Pred. No. 1.7;  
 Matches 10; Conservative 2; Mismatches 3; Indels 0; Gaps 0;  
 Qy 1 PAGPMQPNKGDKGVG 15  
 Db 1623 PPGPVGPRGRDGEGV 1637

RESULT 9

ID CA12_MOUSE	STANDARD;	PRT;	1459 AA.
AC P28481;			
DT 01-DEC-1992 (Rel. 24, Created)			
DT 01-DEC-1992 (Rel. 24, Last annotation update)			
DE Collagen alpha 1(I) chain precursor [Contains: Chondrocalcin].			
GN COL2A1.			
OS Mus musculus (Mouse)			
OC Mammalia; Eutheria; Rodentia; Sciurognath; Muridae; Murinae; Mus.			
NCBI_TaxId=10090;			
[1]			
SEQUENCE FROM N.A.			
RX MEDLINE=911358469; PubMed=1885613;			
RA Netsranta M., Toman D., de Crombrugge B., Vuorio E.;			
RT "Mouse type II collagen gene. Complete nucleotide sequence, exon			
RT structure, and alternative splicing";			
J. Biol. Chem. 266:16662-16669(1991).			
[2]			
SEQUENCE OF 1455-1459 FROM N.A.			
RX MEDLINE=91274355; PubMed=2054384;			
RA Metsranta M., Toman D., de Crombrugge B., Vuorio E.;			
RT "Specific hybridization probes for mouse type I, II, III and IX			
RT collagen mRNAs";			
RL Biochim. Biophys. Acta 1089:241-243(1991).			
CC -I- FUNCTION: COLLAGEN TYPE III IS SPECIFIC FOR CARTILAGINOUS TISSUES.			
CC -I- SUBUNIT: TRIMERS OF IDENTICAL ALPHA 1(II) CHAINS.			
CC -I- PPTM: PROLINES AT THE THIRD POSITION OF THE TRIPETIDE REPEATING			
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.			
CC -I- SIMILARITY: CONTAINS 1 WWC DOMAIN.			

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/or send an email to license@isb-sib.ch>).

DR EMBL; X57982; CAA41047.1; -.  
 DR MGI; MGI:80452; Col2al.  
 DR InterPro; IPR00087; Collagen.  
 DR InterPro; IPR00885; Fib.collagen\_C.  
 DR InterPro; IPR01007; WFEC.  
 DR Pfam; PF01410; COLFI\_1.  
 DR Pfam; PF01391; Collagen; 17.  
 DR Pfam; PF0093; vwc; 1.  
 DR ProDom; PDO02018; Fib.collagen\_C; 1.  
 DR SMART; SM0038; COLEF\_1.  
 DR SMART; SM0214; vwc; 1.  
 DR PROSITE; PS01248; vWFC; 1.  
 DR KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation; Glycoprotein; Collagen; Cartilage; Signal; Alternative splicing.  
 FT SIGNAL 1 25 (BY SIMILARITY); COLLAGEN ALPHA 1(II) CHAIN.  
 FT PROPEP 154 1213 AMINO-TERMINAL PROPEPTIDE  
 FT PROPEP 26 153 CARBOXY-TERMINAL PROPEPTIDE  
 FT CHAIN 154 1213 (CHONDROCALCIN).  
 FT DOMAIN 32 89 VWF\_C.  
 FT DOMAIN 173 1186 TRIPLE-HELICAL REGION.  
 FT DOMAIN 1187 1213 NONHELICAL REGION (C-TERMINAL).  
 FT VARSPIC 29 29 Q -> R (IN SHORT ISOFORM).  
 FT VARSPIC 30 98 MISSING (IN SHORT ISOFORM).  
 SQ SEQUENCE 1459 AA; 139154 MW; FGC84FAT7C522E7F2 CRC64;

Query Match 62.0%; Score 57; DB 1; Length 1459;  
 Best Local Similarity 73.3%; Pred. No. 2.3;  
 Matches 11; Conservative 0; Mismatches 0; Indels 4; Gaps 0;  
 Qy 1 PAGPMQPNKGDKGVG 15  
 Db 888 PAGPPCPAGKDPKG 902

RESULT 10

ID CA12_CHICK	STANDARD;	PRT;	369 AA.
AC P02460;			
DT 21-JUL-1986 (Rel. 01, Created)			
DT 01-NOV-1988 (Rel. 09, Last sequence update)			
DE Collagen alpha 1(II) chain precursor (Fragment).			
GN COL2A1.			
OS Gallus gallus (Chicken)			
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;			
OC Gallus.			
NCBI_TaxId=9031;			
[1]			
SEQUENCE OF 1-193 FROM N.A.			
RX MEDLINE=83306812; PubMed=3840018;			
RA Deak F., Airgraes W.S., Kiss I., Sparks K.J., Goetinck P.F.;			
RT "Primary structure of the telopeptide and a portion of the helical domain of chicken type II procollagen as determined by DNA sequence analysis";			
RL Biochem. J. 229:189-196(1985).			
[2]			
SEQUENCE OF 82-369 FROM N.A.			
RX MEDLINE=84239738; PubMed=6330084;			
RA Sandell L.J., Prentice H.L., Kravits D., Upholt W.B.;			
RT "Structure and sequence of the chicken type II procollagen gene. Characterization of the region encoding the carboxyl-terminal telopeptide and propeptide";			
RT J. Biol. Chem. 259:7826-7834(1984).			
RL [3]			
SEQUENCE OF 114-369 FROM N.A.			
RP Ninomiya Y., Showalter A.M., van der Rest M., Seidah N.G.,			
RA Christen M., Olsen B.R.;			
RT "Structure of the carboxyl propeptide of chicken type II procollagen determined by DNA and protein sequence analysis";			
RT Biochemistry 23:617-624(1984).			

DR EMBL; X57982; CAA41047.1; -.  
 DR MGI; MGI:80452; Col2al.  
 DR InterPro; IPR00087; Collagen.  
 DR InterPro; IPR00885; Fib.collagen\_C.  
 DR InterPro; IPR01007; WFEC.  
 DR Pfam; PF01410; COLFI\_1.  
 DR Pfam; PF01391; Collagen; 17.  
 DR Pfam; PF0093; vwc; 1.  
 DR ProDom; PDO02018; Fib.collagen\_C; 1.  
 DR SMART; SM0038; COLEF\_1.  
 DR SMART; SM0214; vwc; 1.  
 DR PROSITE; PS01248; vWFC; 1.  
 DR KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation; Glycoprotein; Collagen; Cartilage; Signal; Alternative splicing.  
 FT SIGNAL 1 25 (BY SIMILARITY); COLLAGEN ALPHA 1(II) CHAIN.  
 FT PROPEP 154 1213 AMINO-TERMINAL PROPEPTIDE  
 FT PROPEP 26 153 CARBOXY-TERMINAL PROPEPTIDE  
 FT CHAIN 154 1213 (CHONDROCALCIN).  
 FT DOMAIN 32 89 VWF\_C.  
 FT DOMAIN 173 1186 TRIPLE-HELICAL REGION.  
 FT DOMAIN 1187 1213 NONHELICAL REGION (C-TERMINAL).  
 FT VARSPIC 29 29 Q -> R (IN SHORT ISOFORM).  
 FT VARSPIC 30 98 MISSING (IN SHORT ISOFORM).  
 SQ SEQUENCE 1459 AA; 139154 MW; FGC84FAT7C522E7F2 CRC64;

Query Match 62.0%; Score 57; DB 1; Length 1459;  
 Best Local Similarity 73.3%; Pred. No. 2.3;  
 Matches 11; Conservative 0; Mismatches 0; Indels 4; Gaps 0;  
 Qy 1 PAGPMQPNKGDKGVG 15  
 Db 888 PAGPPCPAGKDPKG 902

RESULT 10

ID CA12_CHICK	STANDARD;	PRT;	369 AA.
AC P02460;			
DT 21-JUL-1986 (Rel. 01, Created)			
DT 01-NOV-1988 (Rel. 09, Last sequence update)			
DE Collagen alpha 1(II) chain precursor (Fragment).			
GN COL2A1.			
OS Gallus gallus (Chicken)			
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;			
OC Gallus.			
NCBI_TaxId=9031;			
[1]			
SEQUENCE OF 1-193 FROM N.A.			
RX MEDLINE=83306812; PubMed=3840018;			
RA Deak F., Airgraes W.S., Kiss I., Sparks K.J., Goetinck P.F.;			
RT "Primary structure of the telopeptide and a portion of the helical domain of chicken type II procollagen as determined by DNA sequence analysis";			
RL Biochem. J. 229:189-196(1985).			
[2]			
SEQUENCE OF 82-369 FROM N.A.			
RX MEDLINE=84239738; PubMed=6330084;			
RA Sandell L.J., Prentice H.L., Kravits D., Upholt W.B.;			
RT "Structure and sequence of the chicken type II procollagen gene. Characterization of the region encoding the carboxyl-terminal telopeptide and propeptide";			
RT J. Biol. Chem. 259:7826-7834(1984).			
RL [3]			
SEQUENCE OF 114-369 FROM N.A.			
RP Ninomiya Y., Showalter A.M., van der Rest M., Seidah N.G.,			
RA Christen M., Olsen B.R.;			
RT "Structure of the carboxyl propeptide of chicken type II procollagen determined by DNA and protein sequence analysis";			
RT Biochemistry 23:617-624(1984).			

CC -!- FUNCTION: COLLAGEN TYPE II IS SPECIFIC FOR CARTILAGINOUS TISSUES.  
 CC -!- SUBUNIT: TRIMERS OF IDENTICAL ALPHA 1(II) CHAINS.  
 CC -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPETIDE REPEATING  
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see [http://www.isb-sib.ch/announce/or\\_send\\_an\\_email\\_to\\_license@isb-sib.ch](http://www.isb-sib.ch/announce/or_send_an_email_to_license@isb-sib.ch)).

CC EMBL: X02663; CAA26499\_1; -.  
 DR EMBL: L00063; AAB9967\_1; -.  
 DR EMBL: L00061; AAB9967\_1; JOINED.  
 DR EMBL: L00062; AAB9967\_1; JOINED.  
 PIR: A02860; CGCHC.  
 DR InterPro: IPR00085; Fib\_collagen\_C.  
 DR InterPro: IPR00107; VWF\_C.  
 Pfam: PF01410; COLFI; 1.  
 DR Pfam: PF01391; Collagen; 1.  
 DR Prodrom: PD002078; Fib\_collagen\_C; 1.  
 SMART: SM00038; COLFI; 1.  
 DR PROSITE: PS01208; VWF\_C; PARTIAL.  
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
 Glycoprotein; Cartilage; Collagen.  
 FT NON\_TER 1  
 CHAIN <1 123  
 PROPEP 124 369  
 DOMAIN <1 96  
 FT DOMAIN 97 123  
 FT CARBOHYD 270 270  
 FT DISULFID 275 320  
 SQ SEQUENCE 369 AA; 38889 MW; EF530625B08A3B0 CRC64;

Query Match Score 56; DB 1; Length 369;  
 Best Local Similarity 66.7%; Pred. No. 0.88; 1; Mismatches 4; Indels 0; Gaps 0; Gaps 0;

Oy 1 PAGPWGPNGKDGKV 15  
 | ||| ||| | | |  
 Db 47 PP GPVGPGPSGDGS 61

RESULT 11

CA13\_RAT  
 ID CA13\_RAT STANDARD; PRT; 636 AA.  
 AC P13941; 070604;  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Collagen alpha 1(III) chain (fragment).  
 GN COL1A1.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.  
 OX NCBL\_TaxID-1016;  
 RN [1] SEQUENCE FROM N.A.  
 RX MEDLINE-9414571; Pubmed-8286415;  
 RA Glumoff V., Maekeda J.K., Vuorio E.;  
 RT "Cloning of cDNA for rat pro alpha 1(III) collagen mRNA. Different expression patterns of type I and type III collagen and fibronectin genes in experimental granulation tissue.",  
 RT Biochim. Biophys. Acta 1217:41-48(1994).  
 RN [2] SEQUENCE OF 73-636 FROM N.A.  
 RP STRAIN-Sprague-Dawley; TISSUE-Fibroblast;  
 RC Murtz T., Ellerstrom C., Lindmark C., Christerson C.;  
 RL Submitted (APP-1998) to the EMBL/GenBank/DDBJ databases.  
 [3]

RP SEQUENCE OF 308-482 FROM N.A.  
 RX MEDLINE-88296083; Pubmed-2456904;  
 RA Franklin F.R., Hsu C.-Y.J., Meyers J.C., Lin E., Lyttle C.R., Kom B., Mohn K.;  
 RT "Regulation of alpha 2(II), alpha 1(III), and alpha 2(V) collagen mRNAs by estrogen in the immature rat uterus.",  
 RL DNA 7:347-354 (1988).  
 CC -!- FUNCTION: COLLAGEN TYPE III OCCURS IN MOST SOFT CONNECTIVE TISSUES  
 CC -!- ALONG WITH TYPE I COLLAGEN.  
 CC -!- SUBUNIT: TRIMERS OF IDENTICAL ALPHA 1(III) CHAINS. THE CHAINS ARE  
 CC LINKED TO EACH OTHER BY INTERCHAIN DISULFIDE BONDS. TRIMERS ARE  
 CC ALSO CROSS-LINKED VIA HYDROXYLYSINES.  
 CC EMBL: X0369; CAA9832\_1; -.  
 DR EMBL: A0205395; CAA06510\_1; -.  
 DR EMBL: M21354; AAA0942\_1; -.  
 DR PIR: A29905; A29905.  
 DR InterPro: IPR00085; Fib\_collagen\_C.  
 DR InterPro: IPRO01007; VWF\_C.  
 DR Pfam: PF01410; COLFI; 1.  
 DR Prodrom: PD002078; Fib\_collagen\_C; 1.  
 SMART: SM00038; COLFI; 1.  
 DR PROSITE: PS01208; VWF\_C; PARTIAL.  
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
 Glycoprotein; Cartilage; Collagen.  
 FT NON\_TER 1  
 CHAIN <1 375  
 PROPEP 376 636  
 DOMAIN <1 368  
 FT DOMAIN 369 636  
 FT DISULFID 369 368  
 FT CONFLICT 340 340  
 FT CONFLICT 429 429  
 SQ SEQUENCE 636 AA; 6248159F01D01EE CRC64;

Query Match Score 56; DB 1; Length 636;  
 Best Local Similarity 66.7%; Pred. No. 1.5; 0; Mismatches 5; Indels 0; Gaps 0; Gaps 0;

Oy 1 PAGPWGPNGKDGKV 15  
 | ||| ||| | | |  
 Db 313 PVGPHGPPGKDGS 327

RESULT 12

CA1B\_MOUSE  
 ID CA1B\_MOUSE STANDARD; PRT; 1804 AA.  
 AC Q61245; 0640V7;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Collagen alpha 1(IX) chain precursor.  
 GN COL1A1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
 OX NCBL\_TaxID-10090;  
 RN [1] SEQUENCE FROM N.A.

RC	TISSUE=Embryo;	FT	VARSPLIC	329	413	MISSING (IN SHORT ISOFORM).
RX	MEDLINE=96015067; PubMed=8530046;	FT	VARIANT	189	195	LDRSERS -> SIEVRDR (IN CHONDRODYSPLASIA).
RA	Yoshioka R., Inoguchi K., Khaleduzzaman M., Ninomiya Y.,	FT	VARIANT	196	1804	MISSING (IN CHONDRODYSPLASIA).
RA	Andrikopoulos K., Ramirez F.;	FT	SEQUENCE	1804 AA;	180963 MW;	FEGDB9DDE1219A CRC64;
RT	"Coding sequence and alternative splicing of the mouse alpha 1(XI) collagen gene (Col11a1).";	FT	SEQUENCE OF 181-198 FROM N.A., AND CHONDRODYSPLASIA VARIANT.	[2]		
RL	Genomics 28:337-340(1995).	FT	SEQUENCE OF 181-198 FROM N.A., AND CHONDRODYSPLASIA VARIANT.			
RN		FT	SEQUENCE OF 181-198 FROM N.A., AND CHONDRODYSPLASIA VARIANT.			
RP		FT	SEQUENCE OF 181-198 FROM N.A., AND CHONDRODYSPLASIA VARIANT.			
RC		FT	SEQUENCE OF 181-198 FROM N.A., AND CHONDRODYSPLASIA VARIANT.			
RX	STRAIN=C57BL/6;	FT	SEQUENCE OF 181-198 FROM N.A., AND CHONDRODYSPLASIA VARIANT.			
RA	MEDLINE=95163095; PubMed=7859283;	FT	SEQUENCE OF 181-198 FROM N.A., AND CHONDRODYSPLASIA VARIANT.			
RA	Li Y., Lacerda D.A., Warman M.L., Beier D.R., Yoshioka H.,	FT	SEQUENCE OF 181-198 FROM N.A., AND CHONDRODYSPLASIA VARIANT.			
RA	Ninomiya Y., Oxford J.T., Morris N.P., Andrikopoulos K.,	FT	SEQUENCE OF 181-198 FROM N.A., AND CHONDRODYSPLASIA VARIANT.			
RA	Ramirez F., Wardell B.B., Liffert G.D., Teuscher C., Woodward S.R.,	FT	SEQUENCE OF 181-198 FROM N.A., AND CHONDRODYSPLASIA VARIANT.			
RT	Taylor B.A., Seegmiller B.R., Olsen B.R.;	FT	SEQUENCE OF 181-198 FROM N.A., AND CHONDRODYSPLASIA VARIANT.			
RT	"A fibrillar collagen gene, Col11a1, is essential for skeletal morphogenesis.";	FT	SEQUENCE OF 181-198 FROM N.A., AND CHONDRODYSPLASIA VARIANT.			
RL	Cell 80:423-430(1995).	FT	SEQUENCE OF 181-198 FROM N.A., AND CHONDRODYSPLASIA VARIANT.			
CC	-!- FUNCTION: MAY PLAY AN IMPORTANT ROLE IN FIBRILLOGENESIS BY CONTROLLING LATERAL GROWTH OF COLLAGEN II FIBRILS.	FT	SEQUENCE OF 181-198 FROM N.A., AND CHONDRODYSPLASIA VARIANT.			
CC	-!- SUBUNIT: TRIMERS COMPOSED OF THREE DIFFERENT CHAINS: ALPHA 1(XI)/ALPHA 2(XI) AND ALPHA 3(XI). ALPHA 3(XI) IS A POST-TRANSLATIONAL MODIFICATION OF ALPHA 1(III). ALPHA 1(V) CAN ALSO BE FOUND INSTEAD OF ALPHA 1(III).	FT	SEQUENCE OF 181-198 FROM N.A., AND CHONDRODYSPLASIA VARIANT.			
CC	-!- PIM: PROLINES AT THE THIRD POSITION OF THE TRIPePTIDE REPEATING UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.	FT	SEQUENCE OF 181-198 FROM N.A., AND CHONDRODYSPLASIA VARIANT.			
CC	-!- DISEASE: DEFECTS IN COL11A1 ARE ASSOCIATED WITH CHONDRODYSPLASIA, AN AUTOSOMAL RECESSIVE DISEASE CHARACTERIZED BY SKELETAL DEFECTS CAUSED BY ABNORMALITIES IN THE CARTILAGE OF LIMBS, RIBS, MANDIBLES AND TRACHEA.	FT	SEQUENCE OF 181-198 FROM N.A., AND CHONDRODYSPLASIA VARIANT.			
CC	-!- SIMILARITY: BELONGS TO THE FIBRILLAR CLASS OF COLLAGENS.	FT	SEQUENCE OF 181-198 FROM N.A., AND CHONDRODYSPLASIA VARIANT.			
CC	-!- SIMILARITY: HIGH, TO ALPHA 1(V) AND ALPHA 3(V) CHAINS.	FT	SEQUENCE OF 181-198 FROM N.A., AND CHONDRODYSPLASIA VARIANT.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to license@isb-sib.ch).	FT	SEQUENCE OF 181-198 FROM N.A., AND CHONDRODYSPLASIA VARIANT.			
CC	EMBL; D38162; BRA07367.1. -	FT	SEQUENCE OF 181-198 FROM N.A., AND CHONDRODYSPLASIA VARIANT.			
DR	EMBL; S74574; AAB3439.1. -	FT	SEQUENCE OF 181-198 FROM N.A., AND CHONDRODYSPLASIA VARIANT.			
DR	MGD; MGI:80446; Coll11a1.	FT	SEQUENCE OF 181-198 FROM N.A., AND CHONDRODYSPLASIA VARIANT.			
DR	InterPro; IPR00088; Collagen.	FT	SEQUENCE OF 181-198 FROM N.A., AND CHONDRODYSPLASIA VARIANT.			
DR	InterPro; IPR00088; Fib_collagen_C.	FT	SEQUENCE OF 181-198 FROM N.A., AND CHONDRODYSPLASIA VARIANT.			
DR	InterPro; IPR00191; Laminin_G.	FT	SEQUENCE OF 181-198 FROM N.A., AND CHONDRODYSPLASIA VARIANT.			
DR	InterPro; IPR003129; TSPN.	FT	SEQUENCE OF 181-198 FROM N.A., AND CHONDRODYSPLASIA VARIANT.			
DR	Pfam; PF01410; COLF1; 1.	FT	SEQUENCE OF 181-198 FROM N.A., AND CHONDRODYSPLASIA VARIANT.			
DR	Pfam; PF01391; Collagen; 18.	FT	SEQUENCE OF 181-198 FROM N.A., AND CHONDRODYSPLASIA VARIANT.			
DR	ProDom; PD002078; Fib_collagen_C; 1.	FT	SEQUENCE OF 181-198 FROM N.A., AND CHONDRODYSPLASIA VARIANT.			
DR	SMART; SM00038; COLF1; 1.	FT	SEQUENCE OF 181-198 FROM N.A., AND CHONDRODYSPLASIA VARIANT.			
DR	SMART; SM00282; LamG; 1.	FT	SEQUENCE OF 181-198 FROM N.A., AND CHONDRODYSPLASIA VARIANT.			
DR	SMART; SM00210; TSPN; 1.	FT	SEQUENCE OF 181-198 FROM N.A., AND CHONDRODYSPLASIA VARIANT.			
KW	Extracellular matrix; Collagen; Signal; Alternative splicing; Glycoprotein; Disease mutation.	FT	SEQUENCE OF 181-198 FROM N.A., AND CHONDRODYSPLASIA VARIANT.			
KW	Disease mutation.	FT	SEQUENCE OF 181-198 FROM N.A., AND CHONDRODYSPLASIA VARIANT.			
FT	SIGNAL 1 35 POTENTIAL.	FT	SEQUENCE OF 181-198 FROM N.A., AND CHONDRODYSPLASIA VARIANT.			
FT	PROPEP 36 511 AMINO-TERMINAL PROPEPTIDE (POTENTIAL).	FT	SEQUENCE OF 181-198 FROM N.A., AND CHONDRODYSPLASIA VARIANT.			
FT	CHAIN 512 1561 COLLAGEN ALPHA 1(XI) CHAIN.	FT	SEQUENCE OF 181-198 FROM N.A., AND CHONDRODYSPLASIA VARIANT.			
FT	PROPEP 1562 1804 CARBOXY-TERMINAL PROPEPTIDE.	FT	SEQUENCE OF 181-198 FROM N.A., AND CHONDRODYSPLASIA VARIANT.			
FT	DOMAIN 37 417 NONHELIICAL REGION.	FT	SEQUENCE OF 181-198 FROM N.A., AND CHONDRODYSPLASIA VARIANT.			
FT	DOMAIN 418 506 TRIPLE-HELICAL REGION (INTERRUPTED).	FT	SEQUENCE OF 181-198 FROM N.A., AND CHONDRODYSPLASIA VARIANT.			
FT	DOMAIN 507 509 SHORT NONHELIICAL SEGMENT.	FT	SEQUENCE OF 181-198 FROM N.A., AND CHONDRODYSPLASIA VARIANT.			
FT	DOMAIN 510 510 TELOPEPTIDE.	FT	SEQUENCE OF 181-198 FROM N.A., AND CHONDRODYSPLASIA VARIANT.			
FT	DOMAIN 528 1540 TRIPLE-HELICAL REGION.	FT	SEQUENCE OF 181-198 FROM N.A., AND CHONDRODYSPLASIA VARIANT.			
FT	DOMAIN 1541 1561 NONHELIICAL REGION (C-TERMINAL).	FT	SEQUENCE OF 181-198 FROM N.A., AND CHONDRODYSPLASIA VARIANT.			
FT	CARBONYD 1638 1638 N-LINKED (GLCNAC. . .) (POTENTIAL).	FT	SEQUENCE OF 181-198 FROM N.A., AND CHONDRODYSPLASIA VARIANT.			
FT	SITE 610 610 CROSSLINKING.	FT	SEQUENCE OF 181-198 FROM N.A., AND CHONDRODYSPLASIA VARIANT.			
FT	SITE 1450 1450 CROSSLINKING.	FT	SEQUENCE OF 181-198 FROM N.A., AND CHONDRODYSPLASIA VARIANT.			
FT	VARSPPLIC 329 413 MISSING (IN SHORT ISOFORM).	FT	SEQUENCE OF 181-198 FROM N.A., AND CHONDRODYSPLASIA VARIANT.			
FT	VARIANT 189 195 LDRSERS -> SIEVRDR (IN CHONDRODYSPLASIA).	FT	SEQUENCE OF 181-198 FROM N.A., AND CHONDRODYSPLASIA VARIANT.			
FT	VARIANT 196 1804 MISSING (IN CHONDRODYSPLASIA).	FT	SEQUENCE OF 181-198 FROM N.A., AND CHONDRODYSPLASIA VARIANT.			
FT	SEQUENCE 1804 AA; 180963 MW; FEGDB9DDE1219A CRC64;	FT	SEQUENCE OF 181-198 FROM N.A., AND CHONDRODYSPLASIA VARIANT.			
SO	Query Match Best Local Similarity 66.7%; Score 56; DB 1; Length 1804; Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0;	FT	SEQUENCE OF 181-198 FROM N.A., AND CHONDRODYSPLASIA VARIANT.			
Qy	1 PAGPKGPNGKKGKV 15	FT	SEQUENCE OF 181-198 FROM N.A., AND CHONDRODYSPLASIA VARIANT.			
Db	1224 PRGPQCPNGGQPGQ 1238	FT	SEQUENCE OF 181-198 FROM N.A., AND CHONDRODYSPLASIA VARIANT.			
RESULT 13						
CA1B_HUMAN						
ID CA1B_HUMAN STANDARD; P12107; Q04034; AC P12107; Q04034; DT 01-OCT-1989 (Rel. 12, Created)						
DT 01-NOV-1990 (Rel. 16, Last sequence update)						
DT 16-OCT-2001 (Rel. 40, Last annotation update)						
DE Collagen alpha 1(XI) chain precursor.						
GN Homo sapiens (Human).						
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Eucoleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.						
OX NCBI_TaxID=9606;						
RN [1] J. Biol. Chem. 265:6423-6426(1990).						
RP SEQUENCE OF 538-1806 FROM N.A., AND PARTIAL SEQUENCE.						
RX MEDLINE=89034222; PubMed=3182841; RA Yoshioka H., Ramirez F.; RT "Pro-alpha 1(XI) collagen. Structure of the amino-terminal propeptide and expression of the gene in tumor cell lines.";						
RA Rodriguez M., Yoshioka H., Rodriguez E., van der Rest M., Kimura T., RA Ninomiya Y., Olsen B., Ramirez F.; RT Cloning and sequencing of pro-alpha 1 (XI) collagen cDNA and demonstrates that type XI belongs to the fibrillar class of collagens and reveals that the expression of the gene is not restricted to cartilaginous tissue.";						
RL J. Biol. Chem. 263:17159-17166(1988).						
RN [3] [J. Biol. Chem. 263:17159-17166(1988)].						
RP ALTERNATIVE SPlicing.						
RC TISSUE=Blood;						
FT MEDLINE=95238468; PubMed=7721876; RA Zhdikova N.I., Justice S.R., Mayne R.; RT "Alternative mRNA processing in the variable region of the pro-alpha 1(XI) and pro-alpha 2(XI) collagen chains.";						
RL J. Biol. Chem. 270:9486-9493(1995).						
CC -!- FUNCTION: MAY PLAY AN IMPORTANT ROLE IN FIBRILLOGENESIS BY CONTROLLING LATERAL GROWTH OF COLLAGEN II FIBRILS.						
CC -!- SUBUNIT: TRIMERS COMPOSED OF THREE DIFFERENT CHAINS: ALPHA 1(XI), ALPHA 2(XI) AND ALPHA 3(XI). ALPHA 1(V) CAN ALSO BE FOUND INSTEAD OF ALPHA 3(XI)=-(II).						
CC -!- ALTERNATIVE PRODUCTS: ALTERNATIVE FORMS OF THIS COLLAGEN ARE PRODUCED BY ALTERNATIVE SPlicing. THERE IS ALTERNATIVE USAGE OF EXON 11A OR EXON 11B. BOTH EXONS 11A AND 11B ARE PRESENT IN TRANSCRIPTS FROM CARTILAGE BUT EXON 11B IS PREFERentially UTILIZED IN TRANSCRIPTS FROM TENDON.						
CC -!- TISSUE SPECIFICITY: CARTILAGE, PLACENTA AND SOME TUMOR OR VIRALLY TRANSFORMED CELL LINES.						
CC -!- PIM: PROLINES AT THE THIRD POSITION OF THE TRIPePTIDE REPEATING UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.						
CC -!- SIMILARITY: BELONGS TO THE FIBRILLAR CLASS OF COLLAGENS.						
CC -!- SIMILARITY: HIGH, TO ALPHA 1(V) AND ALPHA 3(V) CHAINS.						
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to license@isb-sib.ch).						

CC	use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to license@isb-sib.ch).
CC	-----
EMBL:	J04177; AAA51891.1; -.
EMBL:	L38956; AAA79171.1; -.
PIR:	A31795; A31795.
MIM:	120280; .
InterPro:	IPR000867; Collagen.
InterPro:	IPR000865; Fib_collagen_C.
InterPro:	IPR001791; laminin_G.
InterPro:	IPR03139; TSPN.
Pfam:	PF01410; COLFI; 1.
Pfam:	PF01391; Collagen; 18.
Pfam:	PF02210; TSN; 1.
ProDom:	PD002078; Fib_collagen_C; 1.
SMART:	SM00338; COLFI; 1.
SMART:	SM00282; Lang; 1.
SMART:	SM00210; TSPN; 1.
KW	Glycoprotein; Collagen; Signal; Alternative splicing.
FT	SIGNAL 1 36 POTENTIAL.
FT	PROPEP 37 511 AMINO-TERMINAL PROPEPTIDE (POTENTIAL).
FT	CHAIN 512 1563 COLLAGEN ALPHA 1(III) CHAIN.
FT	PROPEP 1564 1806 CARBOXYL-TERMINAL PROPEPTIDE.
FT	DOMAIN 37 419 NONHETICAL REGION.
FT	DOMAIN 420 508 TRIPLE-HETICAL REGION (INTERRUPTED).
FT	DOMAIN 509 511 SHORT NONHETICAL SEGMENT.
FT	DOMAIN 512 528 TELEOPETIDE.
FT	DOMAIN 529 1542 TRIPLE-HETICAL REGION.
FT	DOMAIN 1543 1563 NONHETICAL REGION (C-TERMINAL).
FT	CARBOHYD 1640 1640 N-LINKED (GLCNAC. . . ) (POTENTIAL).
FT	SITE 612 612 CROSSLINKING.
FT	VARSPLIC 1452 299 CROSSLINKING.
FT	VARSPLIC 261 299 YAPDIDTYEGEAEKKAESTVGPVTEIAQK -> KKKSNKKMKTIVATSKESKSKFPKSEFKKK SYQSKAKLGVK (IN ISOFORMS WITH EXON IIB).
FT	SEQUENCE 1806 AA: 181137 MW: C27AD86180BCEDF5 CRC64;
Qy	1 PAGPWGPNKGDKGVG 15
Db	1226 PRGPQGPNGADGQG 1240
RESULT 14	
CA13_CHICK	
ID CA13_CHICK STANDARD: PRT: 1262 AA.	
AC P12105; P77958; P79759; 090794; 092029;	
DT 01-OCT-989 (Rel. 12, Created)	
DT 16-OCT-2001 (Rel. 40, Last sequence update)	
DT 01-MAR-2002 (Rel. 41, Last annotation update)	
DE Collagen alpha 1(III) chain precursor (Fragments).	
GN COLJAI.	
OS Gallus gallus (Chicken).	
OC Eukaryota; Chordata; Craniata; Vertebrata; Buteleostomi;	
OC Archosauia; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;	
OC Gallus.	
NCBI_TAXID=9031;	
RN [1]	SEQUENCE OF 1-886 FROM N.A.
RC TISSUE-Kidney;	
RX MEDLINE-94265842; PubMed-8206952;	
RA Nah H.-D., Niu Z., Adams S.L.;	
RT "An alternative transcript of the chick type III collagen gene that does not encode type III collagen.";	
RL J. Biol. Chem. 269:16443-16448(1994).	
RN [2]	SEQUENCE OF 29-96; 332-397; 431-484; 503-535 AND 869-976 FROM N.A.
RP RX MEDLINE-84270696; PubMed-654770;	
RA Tamada Y., Liou G., Mudryj M., Obici S., de Crombrugge B.;	
RT "Conservation of the sizes for one but not another class of exons in two chick collagen genes.";	
RL Nature 310:333-337(1984).	
RN [3]	SEQUENCE OF 977-1262 FROM N.A.
RP RX MEDLINE-83220816; PubMed-6856474;	
RA Yamada Y., Kuhn K., de Crombrugge B.;	
RT "A conserved nucleotide sequence, coding for a segment of the C-propeptide, is found at the same location in different collagen genes.";	
RT RL Nucleic Acids Res. 11:2733-2744(1983).	
CC RT -I FUNCTION: COLLAGEN TYPE III OCCURS IN MOST SOFT CONNECTIVE TISSUES ALONG WITH TYPE I COLLAGEN.	
CC CC -I - SUBUNIT: TRIMERS OF IDENTICAL ALPHA 1(III) CHAINS. THE CHAINS ARE LINKED TO EACH OTHER BY INTERCHAIN DISULFIDE BONDS. TRIMERS ARE ALSO CROSS-LINKED VIA HYDROXYLYSINES.	
CC CC -I - UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.	
CC CC -I - SIMILARITY: CONTAINS 1 VWFC DOMAIN.	
CC CC This SWISS-PROT entry is copyright. It is produced through collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to license@isb-sib.ch).	
CC CC -----	
DR DR EMBL: U07973; AAC03407.1; .	
DR DR EMBL: X00822; CAB52686.1; .	
DR DR EMBL: X00823; CAB52686.1; JOINED.	
DR DR EMBL: X00826; CAB5397.1; ALT_SEQ.	
DR DR EMBL: X00827; CAB5398.1; .	
DR DR EMBL: X00828; CAB5399.1; .	
DR DR EMBL: X00830; CAB52401.1; .	
DR DR EMBL: X00831; CAB52402.1; .	
DR DR EMBL: X02302; AAD15299.1; .	
DR DR EMBL: K02301; AAD15298.1; .	
DR DR EMBL: M36662; AAA18519.1; ALT_SEQ.	
DR DR PIR: A05269; A05269.	
DR DR InterPro: IPR00087; Collagen.	
DR DR InterPro: IPR00085; Fib_collagen_C.	
DR DR InterPro: IPR00107; VWFC.	
DR DR Pfam: PF01391; Collagen; 13.	
DR DR ProDom: PD002078; Fib_collagen_C; 1.	
DR DR SMART: SM0014; VWFC; 1.	
DR DR PROSITE: PS01208; VWFC; 1.	
DR DR KW EXTRACELLULAR MATRIX; Connective tissue; Repeat; Hydroxylation; Glycoprotein; Collagen; Signal; POTENTIAL.	
FT SIGNAL 1 23 AMINO-TERMINAL PROPEPTIDE (BY SIMILARITY).	
FT PROPEP 24 144 COLLAGEN ALPHA 1(III) CHAIN CARBOXYL-TERMINAL PROPEPTIDE (BY SIMILARITY).	
FT CHAIN 145 1003 COLLAGEN ALPHA 1(III) CHAIN CARBOXYL-TERMINAL PROPEPTIDE (BY SIMILARITY).	
FT PROPEP 1004 1262 COLLAGEN ALPHA 1(III) CHAIN CARBOXYL-TERMINAL PROPEPTIDE (BY SIMILARITY).	
FT DOMAIN 229 88 VWFC.	
FT DOMAIN 145 164 NONHETICAL REGION (N-TERMINAL) (BY SIMILARITY).	
FT DOMAIN 165 994 INTERCHAIN (BY SIMILARITY).	
FT DOMAIN 995 995 TRIPLE-HETICAL REGION (BY SIMILARITY).	
FT DOMAIN 262 262 HYDROXYLATION (BY SIMILARITY).	
FT MOD_RES 283 283 HYDROXYLATION (BY SIMILARITY).	

FT MOD\_RES 859 859 HYDROXYLATION (BY SIMILARITY).  
 FT CARBONYD 1163 1163 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CONFLICT 96 96 E -> K (IN REF. 2).  
 FT CONFLICT 1132 1132 F -> S (IN REF. 3).  
 FT SEQUENCE 1262 AA; 121249 MW; 96ABE/B2E9DEB43D CRC64;  
 SQ Query Match 59 %; Score 55; DB 1; Length 1262;  
 Best Local Similarity 66.7%; Pred. No. 3.7; 1; Mismatches 4; Indels 0; Gaps 0;  
 Matches 10; Conservative 1; Mismatches 1; Mismatches 1; Gaps 0;  
 Qy 1 PAGPQPNKGDKGVG 15  
 Db 942 PPGPAGPPGKDGKGG 956

RESULT 15  
 CA21\_CHICK STANDARD; PRT; 1362 AA.  
 ID CA21\_CHICK STANDARD; PRT; 1362 AA.  
 AC P02467; 090795; 090797; P87422; Q90758; Q92014; P87491; Q90792;  
 DT 21-JUL-1996 (Rel. 01, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DE 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Collagen alpha 2(I) chain precursor (Fragments).  
 GN COLLA2.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Phasianinae;  
 OC Gallus.  
 NCBI\_TAXID=9031;  
 RN [1] SEQUENCE OF 1-245; 262-448 AND 466-1362 FROM N.A.  
 RX MEDLINE=6135195;  
 RA Tate V.E., Finer M.H., Boedtker H., Doty P.;  
 RT "Chick pro alpha 2 (I) collagen gene: exon location and coding  
 RL Ann. N.Y. Acad. Sci. 460:85-116(1985).  
 RN [2] SEQUENCE OF 1-89 FROM N.A.  
 RX MEDLINE=613246518; Pubmed=6135195;  
 RA Tate V.E., Finer M.H., Boedtker H., Doty P.;  
 RT "Chick pro alpha 2 (I) collagen gene: exon location and coding  
 RL Nucleic Acids Res. 11:91-104(1983).  
 RN [3] SEQUENCE OF 1-14 FROM N.A.  
 RX MEDLINE=6135195;  
 RA Vogeli G., Ohkubo H., Sobel M.E., Yamada Y., Pastan I.,  
 RT "The structure of the chicken alpha 2 collagen gene.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 78:5334-5338(1981).  
 RN [4] SEQUENCE OF 1-33 FROM N.A.  
 RX MEDLINE=6472717; Pubmed=6473103;  
 RA Aho S., Tate V.E., Boedtker H.;  
 RT "Structure of the promoter for chicken alpha 2 type I collagen gene.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 78:5334-5338(1981).  
 RN [5] SEQUENCE OF 1-33 FROM N.A.  
 RX MEDLINE=6472717; Pubmed=6473103;  
 RA Aho S., Tate V.E., Boedtker H.;  
 RT "Location of the 11 bp exon in the chicken pro alpha 2(I) collagen  
 gene.";  
 RL Nucleic Acids Res. 12:6117-6125(1984).  
 RN [6] SEQUENCE OF 1-79 FROM N.A.  
 RX MEDLINE=3678834;  
 RA Finer M.H., Boedtker H., Doty P.;  
 RT "Construction and characterization of cDNA clones encoding the 5' end  
 RT of the chicken pro alpha 1(I) collagen mRNA.";  
 RL Gene 56:71-78(1987).  
 RN [6] SEQUENCE OF 78-92.  
 RC TISSUE-Skin;  
 RX MEDLINE=71115216; Pubmed=5544653;  
 RA Highberger H., Kang A.H., Gross J.;  
 RT "Comparative studies on the amino acid sequence of the alpha 2-CB2  
 RT peptides from chick and rat skin collagens.";  
 RL Biochemistry 10:610-616(1971).  
 RN [7] SEQUENCE OF 74-91; 263-448 AND 1088-1169 FROM N.A.

RX MEDLINE=82050801; PubMed=6272119;  
 RA Wozney J., Hanahan D., Tate V.E., Boedtker H., Doty P.,  
 RT "Structure of the pro alpha 2 (I) collagen gene.";  
 RL Nature 294:129-135(1981).  
 SQ [8] SEQUENCE OF 78-92.  
 RP TISSUE=Skin;  
 RX MEDLINE=70131186; PubMed=4313735;  
 RA Kang A.H., Gross J.;  
 RT "Amino acid sequence of cyanogen bromide peptides from the amino-  
 terminal region of chick skicollagen.";  
 RL Biochemistry 9:796-804(1970).  
 RN [9] SEQUENCE OF 78-92 AND 415-448.  
 RP TISSUE=Skin;  
 RX MEDLINE=69285369; PubMed=5809220;  
 RA Kang A.H., Igashiki S., Gross J.;  
 RT "Characterization of the cyanogen bromide peptides from the alpha-2  
 chain of chick skin collagen.";  
 RL Biochemistry 8:3200-3204(1969).  
 RN [10] SEQUENCE OF 78-92 AND 415-448.  
 RP TISSUE=Bone;  
 RX MEDLINE=6926882; PubMed=5785233;  
 RA Lane J.M., Miller E.J.;  
 RT "Isolation and characterization of the peptides derived from the  
 alpha 2 chain of chick bone collagen after cyanogen bromide  
 cleavage.";  
 RL Biochemistry 8:2134-2139(1969).  
 RN [11] SEQUENCE OF 566-587 FROM N.A.  
 RX MEDLINE=9074829; Pubmed=364479;  
 RA Lehrach H., Frischaut A.-M., Hanahan D., Wozney J., Fuller F.,  
 RA Crkvenjakov R., Boedtker H., Doty P.;  
 RT "Construction and characterization of a 2.5-kilobase procollagen  
 clone.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 75:5417-5421(1978).  
 RN [12] SEQUENCE OF 902-1362 FROM N.A.  
 RX MEDLINE=81160715; PubMed=6927845;  
 RA Fuller F., Boedtker H.;  
 RT "Sequence determination and analysis of the 3' region of chicken pro-  
 alpha 1(I) and pro-alpha 2(I) collagen messenger ribonucleic acids  
 including the carboxy-terminal propeptide sequences.";  
 RL Biochemistry 20:996-1006(1981).  
 RN [13] SEQUENCE OF 998-1169 AND 1234-1362 FROM N.A.  
 RX MEDLINE=81264246; PubMed=6267043;  
 RA Dickson L.A., Ninomaya Y., Bernard M.P., Pesciotta D.M., Parsons J.,  
 RA Green G., Eikenberry E.F., de Crombrughe B., Vogeli G., Pastan I.,  
 RA Fletzek P., Olsen B.R.;  
 RT "The exon/intron structure of the 3'-region of the pro alpha 2(I)  
 collagen gene.";  
 RL J. Biol. Chem. 256:8407-8415(1981).  
 RN [14] SEQUENCE OF 932-954 AND 968-980 FROM N.A.  
 RX MEDLINE=81064671; Pubmed=6159982;  
 RA Avvedimento V.E., Vogeli G., Yamada Y., Maizel J.V. Jr., Pastan I.,  
 RA de Crombrughe B.;  
 RT "Correlation between splicing sites within an intron and their  
 RT sequence complementarity with u1 RNA.";  
 RL Cell 21:689-696(1980).  
 RN [15] SEQUENCE OF 1266-161; 467-517 AND 926-954 FROM N.A.  
 RX MEDLINE=81112157; PubMed=7460017;  
 RA Iran M., Pastan I., de Crombrughe B.;  
 RA Yamada Y., Avvedimento V.E., Mudryj M., Ohkubo H., Vogeli G.,  
 RA "The collagen gene: evidence for its evolutionary assembly by  
 RT amplification of a DNA segment containing an exon of 54 bp.";  
 RL Cell 22:887-892(1980).  
 CC -1- FUNCTION: TYPE I COLLAGEN IS A MEMBER OF GROUP I COLLAGEN  
 CC (FIBRILLAR FORMING COLLAGEN).  
 CC -1- SUBUNIT: TRIMERS OF ONE ALPHA 2 (I) AND TWO ALPHA 1 (I) CHAINS.

Best Local Similarity 66.7%; Pred. No. 4; Indels 0; Gaps 0	
Matches 10; Conservative 1; Mismatches 4;	Db 1053 PPGPSGPGKGGRNG 1067
RESULT 16	
CAL2_HUMAN	
CA12_HUMAN	
STANDARD;	
PRT; 1418 AA.	
ID CA12_HUMAN	
AC P02458;	
DT 21-JUL-1986 (Rel. 01, Created)	
DT 01-JAN-1990 (Rel. 13, Last sequence update)	
DT 16-OCT-2001 (Rel. 40, Last annotation update)	
DE Collagen alpha 1(II) chain precursor [Contains: Chondrocalcin].	
GN COL2A1.	
OS Homo sapiens (Human).	
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
OX NCBI_Taxid=9606;	
RN [1]	
RP SEQUENCE FROM N.A.	
RX MEDLINE-#90067946; PubMed=2587267;	
RX RA Su M.W., Lee B., Ramirez F., Machado M., Horton W.;	
RT RT "Nucleotide sequence of the full length cDNA encoding for human type	
RT II procollagen";	
RL Nucleic Acids Res. 17:9473-9473(1989).	
RN [2]	
RP SEQUENCE OF 1-28 FROM N.A.	
RX MEDLINE-#90067946; PubMed=3021582;	
RX RA Nunez A.M., Kohno K., Martin G.R., Yamada Y.;	
RL Submitted (DEC-1988) to the EMBL/GenBank/DDBJ databases.	
RN [4]	
RP SEQUENCE OF 963-1418 FROM N.A.	
RX MEDLINE-#85190524; PubMed=385798;	
RA Cheah K.S.E., Stoker N.G., Griffin J.R., Grosvenor F.G., Solomon E.;	
RT RT "Identification and characterization of the human type II collagen	
RT gene (COL2A1).";	
RL Proc. Natl. Acad. Sci. U.S.A. 82:2555-2559(1985).	
RN [5]	
RP SEQUENCE OF 1120-1398 FROM N.A.	
RX MEDLINE-#85306861; PubMed=3840017;	
RA Elima K., Maekelae J.K., Vuorio T., Kauppinen S., Knowles J.,	
RA Vuorio E.;	
RT "Construction and identification of a cDNA clone for human type II	
RT procollagen mRNA.";	
RT Biochem. J. 228:183-188(1985).	
RL [6]	
RP SEQUENCE OF 1106-1418 FROM N.A.	
RX MEDLINE-#88067771; PubMed=2825137;	
RA Elima K., Vuorio T., Vuorio E.;	
RT "Determination of the single polyadenylation site of the human pro	
RT alpha 1(II) collagen gene.>";	
RL Nucleic Acids Res. 15:9499-9504(1987).	
RN [7]	
RP SEQUENCE OF 1227-1289 FROM N.A.	
RX MEDLINE-#85104139; PubMed=302437;	
RA Nunez A.M., Francomano C.C., Young M.F., Martin G.R., Yamada Y.;	
RT "Isolation and partial characterization of genomic clones coding for	
RT a human pro-alpha 1 (II) collagen chain and demonstration of	
RT restriction fragment length polymorphism at the 3' end of the gene.";	
RL Biochemistry 24:6343-6347(1985).	
RN [8]	
RP SEQUENCE OF 1176-1226 FROM N.A.	
RX MEDLINE-#84118778; PubMed=6320112;	

- RT the human type II procollagen gene.";  
 RL Nucleic Acids Res. 12:1025-1038(1984).  
 RN [9]
- RP SEQUENCE OF 35-167 FROM N.A.  
 RX MEDLINE=9233138; PubMed=2714801;  
 RA Su M.W., Benson-Chanda V., Vassing H., Ramirez F.;  
 RT "Organization of the exons coding for pro alpha 1(II) collagen N-  
 peptide confirms a distinct evolutionary history of this domain of  
 the fibrillar collagen genes";  
 RT Genomics 4:438-441(1989).  
 RL [10]
- RN REVIEW ON VARIANTS.  
 RP MEDLINE=91184577; PubMed=2010058;  
 RX Kuijaniemi H., Thomp G., Prockop D.J.;  
 RA "Mutations in collagen genes: causes of rare and some common diseases  
 in humans.";  
 RT FASEB J. 5:2052-2060(1991).  
 RL [11]
- RN REVIEW ON VARIANTS.  
 RP MEDLINE=97253595; PubMed=9101290;  
 RX Kuijaniemi H., Thomp G., Prockop D.J.;  
 RA "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-  
 associated collagen (type IX), and network-forming collagen (type X)  
 cause a spectrum of diseases of bone, cartilage, and blood vessels.";  
 RT Hum. Mutat. 9:300-315(1997).  
 RL [12]
- RN VARIANT SER-1074.  
 RP MEDLINE=0036509; PubMed=2577591;  
 RX Vissing H., D'Alessio M., Lee B., Ramirez F., Godfrey M.,  
 RA Hollister D.W.;  
 RT "Glycine 1074 to serine substitution in the triple helical domain of pro-  
 alpha 1 (III) collagen results in a lethal perinatal form of short-  
 limbed dwarfism.";  
 RT J. Biol. Chem. 264:18265-18267(1989).  
 RL [13]
- RN VARIANT SEDC 1095-GLY-TYR-1330 DEL.  
 RX MEDLINE=9266907; PubMed=2545071;  
 RA Lee B., Vissing H., Ramirez F., Rogers D., Rimoin D.;  
 RT "Identification of the molecular defect in a family with  
 spondyloepiphyseal dysplasia";  
 RL Science 244:978-980(1989).  
 RN [14]
- RX VARIANT OSTEOARTHRITIS CYS-650.  
 RP MEDLINE=00370826; PubMed=1975693;  
 RA Ala-Kokko L., Baldwin C.T., Moskowitz R.W., Prockop D.J.;  
 RT "Single base mutation in the type II procollagen gene (COL2A1) as a  
 cause of primary osteoarthritis associated with a mild  
 chondrodysplasia.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 87:6565-6568(1990).  
 RL [15]
- RN VARIANT OI-IV VAL-717.  
 RP MEDLINE=0129136; PubMed=2064612;  
 RA Bateman J.F., Hannigan M., Chan D., Cole W.G.;  
 RT "Characterization of a type I collagen alpha 2(I) glycine-586 to  
 valine substitution in osteogenesis imperfecta type IV: Detection of  
 the mutation and prenatal diagnosis by a chemical cleavage method.";  
 RL Biochem. J. 276:765-770(1991).  
 RN [16]
- RN VARIANT OSTEOARTHRITIS CYS-650.  
 RX MEDLINE=91086471; PubMed=1985108;  
 RA Eyre D.R., Weis M.A., Moskowitz R.W.;  
 RT "Cartilage expression of a type II collagen mutation in an inherited  
 form of osteoarthritis associated with a mild chondrodysplasia.";  
 RL J. Clin. Invest. 87:357-361(1991).  
 RN [17]
- RN VARIANT HYPOCHONDROGENESIS GLU-984.  
 RX MEDLINE=93054348; PubMed=1428602;  
 RA Bogert R., Tiller G.E., Wied M.A., Gruber H.E., Rimoin D.L.,  
 RA Conn D.H., Eyre D.R.;  
 RT "An amino acid substitution (Gly883-->Glu) in the collagen alpha  
 J. Biol. Chem. 267:22522-22526(1992).  
 RL [18]
- 
- RP VARIANT HYPOCHONDROGENESIS SSR-705.  
 RX MEDLINE=92262484; PubMed=1374906;  
 RA Horiton W.A., Machado M.A., Ellard J., Campbell D., Bartley J.,  
 Ramirez F., Vitale E., Lee B.;  
 RT "Characterization of a type II collagen gene (COL2A1) mutation  
 identified in cultured chondrocytes from human hypochondrogenesis.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 89:4583-4587(1992).  
 RL [19]
- RN VARIANT WS-II ASP-198.  
 RP MEDLINE=93304428; PubMed=8317498;  
 RX Koirikko J., Rittvaniemi P., Haataja L., Kaaerlaeinen H.,  
 RA Koirikko K.I., Prockop D.J., Ala-Kokko L.;  
 RT "Mutation in type II procollagen (COL2A1) that substitutes aspartate  
 for glycine alpha 1-67 and that causes cataracts and retinal  
 detachment: evidence for molecular heterogeneity in the Wagner  
 syndrome and the Stickler syndrome (arthro-opthalmopathy)." ;  
 RL Ann. J. Hum. Genet. 53:55-61(1993).  
 RN [20]
- RP VARIANT SEMD CYS-840.  
 RX Tiller G.E., Weis M.A., Lachman R.S., Cohn D.H., Rimoin D.L.,  
 RA Eyre D.R.;  
 RT "A dominant mutation in the type II collagen gene (COL2A1) produces  
 spodyloepiphyseal dysplasia (SEMD), Strudwick type.";  
 RL Ann. J. Hum. Genet. 53:A209-A209(1993).
- RP VARIANT OSTEOPHTHALMOPATHY CYS-650.  
 RX MEDLINE=93282819; PubMed=8501190;  
 RA Holderbaum D., Malenud C.J., Moskowitz R.W., Hagg D.M.;  
 RT "Human cartilage from late stage familial osteoarthritis transcribes  
 type II collagen mRNA encoding a cysteine in position 519.";  
 RL Biochem. Biophys. Res. Commun. 192:1169-1174(1993).  
 RP VARIANT SEMD ARG-285.  
 RX MEDLINE=93252400; PubMed=8486375;  
 RA Vikkula M., Rittvaniemi P., Vuorio A.F., Kaitila I., Ala-Kokko L.,  
 RA Peltonen L.; Rittvaniemi P., Vuorio A.F., Kaitila I., Ala-Kokko L.,  
 RT "A mutation in the amino-terminal end of the triple helix of type II  
 collagen causing severe osteochondrodysplasia.";  
 RL Genomics 16:282-285(1993).  
 RN [23]
- RP VARIANT SEDC CYS-206.  
 RX MEDLINE=94063862; PubMed=8244341;  
 RA Williamson C.J., Considine E.L., Knowlton R.G., Reginato A., Neumann G.,  
 RA Harrison D., Buxton P., Jimenez S.A., Prockop D.J.;  
 RT "Spodyloepiphyseal dysplasia and precocious osteoarthritis in a  
 family with an Arg5/5-->Cys mutation in the procollagen type II gene  
 (COL2A1)";  
 RT Hum. Genet. 92:499-505(1993).  
 RL [24]
- RP VARIANT SEDC CYS-920.  
 RX MEDLINE=93315508; PubMed=8325895;  
 RA Chan D., Taylor T.K.F., Cole W.G.;  
 RT "Characterization of an arginine 789 to cysteine substitution in  
 alpha 1 (II) collagen chains of a patient with spodyloepiphyseal  
 dysplasia.";  
 RL J. Biol. Chem. 268:15238-15245(1993).  
 RN [25]
- RP VARIANT SEDC SSR-1128.  
 RX MEDLINE=93140139; PubMed=8423604;  
 RA Cole W.G., Hall R.K., Rogers J.G.;  
 RT "The clinical features of spodyloepiphyseal dysplasia congenita  
 resulting from the substitution of glycine 997 by serine in the alpha  
 1(II) chain of type II collagen.";  
 RT J. Med. Genet. 30:27-35(1993).  
 RL
- |                       |       |       |           |     |    |   |        |      |
|-----------------------|-------|-------|-----------|-----|----|---|--------|------|
| Query                 | Match | 59 8% | Score     | 55  | DB | 1 | Length | 1418 |
| Best Local Similarity |       | 66 7% | Pred. No. | 4.2 |    |   |        |      |
| Matches               | 10    |       | Indels    | 4   |    |   |        |      |
| Conservative          |       |       | Gaps      | 0   |    |   |        |      |
| Mismatches            | 1     |       |           |     |    |   |        |      |
- Qy 1 PAGPWPGNGDGKVG 15.
- Db 1096 PPGPVGPSSKGDKGANG 1110

RESULT 17  
 CALL\_CHICK STANDARD PRN; 1453 AA.  
 ID CALL\_CHICK STANDARD PRN; 1453 AA.  
 AC P02457;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 01-OCT-1988 (Rel. 12, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Collagen alpha 1(I) chain precursor.  
 GN COL1A1.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosaria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 RN NCBI\_TAXID=9031;  
 RP SEQUENCE OF 1-153 FROM N.A.  
 RX MEDLINE=88056316; Pubmed=3678834;  
 RA Finer M.H., Boedtke H., Doty P.;  
 RT "Construction and characterization of cDNA clones encoding the 5' end  
 of the chicken pro-alpha 1(I) collagen mRNA.",  
 RL Gene 56:71-78(1987).  
 RN [2]  
 RP SEQUENCE OF 1-144 FROM N.A.  
 RX MEDLINE=88007542; Pubmed=2830966;  
 RA Finer M.H., Aho S., Gerstenfeld L.C., Boedtke H., Doty P.;  
 RT "Unusual DNA sequences located within the promoter region and the  
 first intron of the chicken pro-alpha 1(I) collagen gene.",  
 RL J. Biol. Chem. 262:13323-13332(1987).  
 RN [3]  
 RP SEQUENCE OF 152-187.  
 RX MEDLINE=82231995; Pubmed=7933229;  
 RA Highberger J.H., Corbett H., Dixit S.N., Yu W., Seyer J.M.,  
 RA Kang A.H., Gross J.;  
 RT "Amino acid sequence of chick skin collagen alpha 1(I)-CB8 and the  
 complete primary structure of the helical portion of the chick skin  
 collagen alpha 1(I) chain.",  
 RL Biochemistry 21:2048-2055(1982).  
 RN [4]  
 RP SEQUENCE OF 1200-1205.  
 RX MEDLINE=72243016; Pubmed=5047697;  
 RA Eye D.R., Glimcher M.J.;  
 RT "Evidence for a previously undetected sequence at the carboxyterminus  
 of the alpha 1 chain of chicken bone collagen.",  
 RL Biochem. Biophys. Res. Commun. 48:720-726(1972).  
 RN [5]  
 RP SEQUENCE OF 981-1453 FROM N.A.  
 RX MEDLINE=81160715; Pubmed=6927845;  
 RA Fuller F., Boedtke H.;  
 RT "Sequence determination and analysis of the 3' region of chicken pro-  
 alpha 1(I) and pro-alpha 2(I) collagen messenger ribonucleic acids  
 including the carboxy-terminal propeptide sequences.",  
 RL Biochemistry 20:996-1006(1981).  
 RN [6]  
 RP SEQUENCE OF 1311-1453 FROM N.A.  
 RX MEDLINE=80134546; Pubmed=6907088;  
 RA Showalter A.M., Pesciotta D.M., Eikenberry E.F., Yamamoto T.,  
 RA Pastan I., Decrombrughe B., Fietzke P.P., Olsen B.R.;  
 RT "Nucleotide sequence of a collagen cDNA-fragment coding for the  
 carboxyl end of pro alpha 1(I)-chains.",  
 RL FEBS Lett. 111:61-65(1980).  
 CC -I- FUNCTION: TYPE I COLLAGEN IS A MEMBER OF GROUP I COLLAGEN  
 CC -I- FUNCTION: FIBRILLAR FORMING COLLAGEN.  
 CC -I- SUBUNIT: TRIMERS OF ONE ALPHA 2(I) AND TWO ALPHA 1(I) CHAINS.  
 CC -I- TISSUE SPECIFICITY: FORMS THE FIBRILS OF TENDON, LIGAMENTS AND  
 CC BONES. IN BONES THE FIBRILS ARE MINERALIZED WITH CALCIUM  
 CC HYDROXYAPATITE.  
 CC -I- PROLINES AT THE THIRD POSITION OF THE TRipeptIDE REPEATING  
 CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
 CC -I- SIMILARITY: CONTAINS 1 WFPC DOMAIN.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to license@isb-sib.ch).  
 CC  
 DR EMBL; M17839; AAA48704.1; JOINED.  
 DR EMBL; M17838; AAA48704.1; JOINED.  
 DR EMBL; V00401; CAM23695.1; -.  
 DR EMBL; M10571; AAA8671.1; ALT\_SEQ.  
 DR EMBL; M17607; AAA48672.1; -.  
 DR PIR; A02857; CGCHIS.  
 DR PIR; A27179; A27179.  
 DR A29367; A29367.  
 DR InterPro; IPR000087; Collagen.  
 DR InterPro; IPR001007; WFPC.  
 DR InterPro; IPR001007; WFPC.  
 DR Pfam; PF01391; Collagen; 18.  
 DR Pfam; PF00093; vwc; 1.  
 DR Prodrom; PD002078; Fib\_colagen\_C; 1.  
 DR SMART; SM00338; COLFI; 1.  
 DR SMART; SM00214; VWC; 1.  
 DR PROSITE; PS01208; WFPC; 1.  
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
 KW Glycoprotein; Collagen; Signal.  
 FT SIGNAL 1 22  
 FT PROPEP 23 151  
 FT CHAIN 152 1205  
 FT PROPEP 1 453  
 FT DOMAIN 331 89  
 FT MOD\_RES 152 152  
 FT PROTEIN 254 254  
 FT MOD\_RES 851 851  
 FT MOD\_RES 1081 1081  
 FT MOD\_RES 1097 1097  
 FT MOD\_RES 1153 1153  
 FT CONFLICT 1187 1187  
 FT CONFLICT 1441 1441  
 FT CONFLICT 1441 1441  
 SQ QY 1 PACPPWGPNGKDDGKG  
 DB 211 PAGPPGKNGDDGEG 15  
 RESULT 18  
 CA13\_MOUSE STANDARD PRN; 1464 AA.  
 ID CA13\_MOUSE STANDARD PRN; 1464 AA.  
 AC P0121; Q61429; Q9CRN7; Created  
 DT 01-AUG-1988 (Rel. 08, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Collagen alpha 1(I,II) chain precursor.  
 GN COL3A1.  
 OS Mus musculus (Mouse).  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OC NCBITAXID=10090;  
 RN RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL X DBA; TISSUE=Embryo;  
 RX MEDLINE=95011609; Pubmed=7936795;  
 RA Tomai D., de Crombrughe B.;  
 RT "The mouse type-III procollagen encoding gene: genomic cloning and  
 RT complete DNA sequence.",  
 RL Gene 147:161-168(1994).

[2] SEQUENCE OF 1-488 FROM N.A.  
 RP MEDLINE=88131189; PubMed=3443309;  
 RX RA Wood L., Theriault N., Vogeli G.;  
 RT "Complete nucleotide sequence of the N-terminal domains of the murine  
 alpha-1 type-III collagen chain.";  
 RT alpha-1 type-III collagen chain.;  
 RL Gene 61:225-230(1987).  
 RN [3]  
 RP SEQUENCE OF 1-28 FROM N.A.  
 RX MEDLINE=88131189; PubMed=3443309;  
 RT "Identification of the promoter and first exon of the mouse alpha 1  
 (III) collagen gene.;"  
 RLT J. Biol. Chem. 260:3773-3777(1985).  
 RN [4]  
 RP SEQUENCE OF 810-1464 FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=embryonic head;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RT (III) collagen gene.;  
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ichii Y.,  
 Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 Alzawa K., Izaka M., Nishi K., Kiyosawa H., Kondo S., Yamakawa I.,  
 Saito T., Okazaki Y., Gojobori T., Bono H., Kasuwa T., Saito R.,  
 Kadota T., Matsuki H.A., Ashburner M., Batalov S., Casavant T.,  
 Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
 Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
 Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
 Brownstein M.J., Built C., Fletcher C., Fujita M., Gariboldi M.,  
 Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
 Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
 Nordone P., Ringwald M., Sakamoto N.,  
 Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,  
 Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,  
 Hayashizaki Y.;  
 RT "Functional annotation of a full-length mouse cDNA collection.;"  
 RL Nature 409:685-690(2001).  
 RN [5]  
 RP SEQUENCE OF 1442-1464 FROM N.A.  
 RC STRAIN=C57BL/  
 RX MEDLINE=91274355; PubMed=205384;  
 RT Metzgeranta M., Toman D., de Crombrugghe B., Vuorio E.;  
 "Specific hybridization probes for mouse type I, II, III and IX  
 collagen mRNAs.;"  
 RT Biochim. Biophys. Acta 1089:241-243(1991).  
 CC -I- FUNCTION: COLLAGEN TYPE III OCCURS IN MOST SOFT CONNECTIVE TISSUES  
 ALONG WITH TYPE I COLLAGEN.  
 -I- SUBUNIT: TRIMERS OF IDENTICAL ALPHA 1(III) CHAINS. THE CHAINS ARE  
 LINKED TO EACH OTHER BY INTERCHAIN DISULFIDE BONDS. TRIMERS ARE  
 ALSO CROSS-LINKED VIA HYDROLYSSES.  
 -I- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPETIDE REPEATING  
 UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
 O-LINKED GLYCAN CONSISTS OF GLC-GAL DISACCHARIDE (BY SIMILARITY).  
 -I- SIMILARITY: CONTAINS 1 VWC DOMAIN.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 or send an email to license@isb-sib.ch).  
 CC EMBL: X52046; CAA36279.1; -.  
 DR EMBL: M10933; AAA37338.1; -.  
 DR EMBL: AK019448; BAB31724.1; -.  
 DR EMBL: X57983; CAA41048.1; -.  
 DR PIR: A22287; A22287.  
 DR PIR: A27353; A27353.  
 DR PIR: S16733; S16373.  
 DR MGD; MGI:88453; Col3al.

DR IntePro: IPR000087; Collagen.  
 DR IntePro: IPR0085; Fib\_collagen\_C.  
 DR InterPro: IPR01007; VWFC.  
 DR Pfam: PF01410; COLFI\_1.  
 DR Pfam: PF01391; Collagen\_17.  
 DR ProDom: PD002078; Fib\_collagen\_C; 1.  
 DR SMART: SM00038; COLFI\_1.  
 DR PROSITE: PS0128; VWFC; 1.  
 KW Glycoprotein; collagen; Signal.  
 FT SIGNAL 1 23 BY SIMILARITY.  
 FT PROPEP 24 154 AMINO-TERMINAL PROPEPTIDE.  
 FT CHAIN 155 1203 COLLAGEN ALPHA 1(III) CHAIN.  
 FT PROPEP 1204 1464 CARBOYL-TERMINAL PROPEPTIDE.  
 FT DOMAIN 31 90 VWFC.  
 FT DOMAIN 155 169 NON-HELICAL REGION (N-TERMINAL).  
 FT DOMAIN 170 1195 TRIPLE-HELICAL REGION.  
 FT DOMAIN 1195 1464 NON-HELICAL REGION (C-TERMINAL).  
 FT CARBOHYD 262 262 O-LINKED (GAL, .) (BY SIMILARITY).  
 FT MOD\_RES 262 262 HYDROXYLATION (BY SIMILARITY).  
 FT MOD\_RES 283 283 HYDROXYLATION (BY SIMILARITY).  
 FT MOD\_RES 859 859 HYDROXYLATION (BY SIMILARITY).  
 FT MOD\_RES 976 976 HYDROXYLATION (BY SIMILARITY).  
 FT MOD\_RES 1093 1093 HYDROXYLATION (BY SIMILARITY).  
 FT MOD\_RES 1105 1105 HYDROXYLATION (BY SIMILARITY).  
 FT DISUFID 1195 1195 INTERCHAIN (BY SIMILARITY).  
 FT DISUFID 1196 1196 INTERCHAIN (BY SIMILARITY);  
 SQ SEQUENCE 1464 AA: 138944 MW: 2104EC7A8066090B CRC64;  
 Qy 1 PAGPMPNGKGKVG 15 Query Match 59.8%; Score 55; DB 1; Length 1464;  
 Db 1140 PVGPHGGKGDTSG 1154 Best Local Similarity 66.7%; Pred. No. 4.3; Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

RESULT 19  
 CA2B\_MOUSE  
 ID CA2B\_MOUSE STANDARD; PRT; 1650 AA.  
 AC 064739; QZ21W0;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Collagen alpha 2(XI) chain precursor.  
 GN COL11A2.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_Taxid=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=129/SVJ;  
 RC Rowen L., Oin S., Madan A., Loretz C., James R., Dors M., Mix L.,  
 RA Hall J., Lasky S., Hood L.;  
 RT "Sequence of the mouse major histocompatibility locus class II  
 region.;"  
 RT Submitted (Oct-1998) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE OF 1-1592 FROM N.A.  
 RC STRAIN=FVB/N AND 129SV; TISSUE=Cartilage;  
 RX MEDLINE=97135795; PubMed=898332;  
 RA Vandenberg P., Vuoristo M.M., Ala-Kokko L., Prockop D.J.;  
 RT "The mouse coll11a2 gene. Some transcripts from the adjacent rxr-beta  
 gene extend into the coll11a2 gene.;"  
 RT Matrix Biol. 15:359-367(1996).  
 CC -I- FUNCTION: MAY PLAY AN IMPORTANT ROLE IN FIBRILLOGENESIS BY  
 CC CONTROLLING LATERAL GROWTH OF COLLAGEN II FIBRILS (BY SIMILARITY).  
 CC -I- SUBUNIT: TRIMERS COMPOSED OF THREE DIFFERENT CHAINS: ALPHA 1(XI),  
 CC ALPHA 2(XI), AND ALPHA 3(XI). ALPHA 3(XI) IS A POST-TRANSLATIONAL  
 CC MODIFICATION OF ALPHA 1(III). ALPHA 1(V) CAN ALSO BE FOUND INSTEAD



RL  
 CC  
 CC  
 CC  
 CC  
 -!- FUNCTION: MAY PLAY AN IMPORTANT ROLE IN FIBRILlogenesis BY CONTROLLING LATERAL GROWTH OF COLLAGEN II FIBRILS.  
 -!- SUBUNIT: TRIMERS COMPOSED OF THREE DIFFERENT CHAINS: ALPHA 1(XI), ALPHA 2(XI), AND ALPHA 3(XI). ALPHA 3(XI) IS A POST-TRANSLATIONAL MODIFICATION OF ALPHA 1(II). ALPHA 1(V) CAN ALSO BE FOUND INSTEAD OF ALPHA 3(XI)=1(II).  
 -!- ALTERNATIVE PRODUCTS: ALTERNATIVE FORMS OF THIS COLLAGEN ARE PRODUCED BY ALTERNATIVE SPLICING. THESE FORMS DIFFER IN THE PRESENCE OR ABSENCE OF THREE SEPARATE REGIONS IN THE N-TERMINAL NON-COLLAGENOUS DOMAIN.  
 -!- PTM: PROLINES AT THE THIRD POSITION OF THE TRipeptIDE REPEATING UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
 -!- PTM: A DISULFIDE-BONDED PEPTIDE CALLED PROLINE/ARGININE-RICH PROTEIN OR PARP IS RELEASED FROM THE AMINO TERMINUS DURING EXTRACELLULAR PROCESSING AND IS SUBSEQUENTLY RETAINED IN THE CARTILAGE MATRIX FROM WHICH IT CAN BE ISOLATED IN SIGNIFICANT AMOUNTS.

-!- DISEASE: DEFECTS IN COL1A2 ARE THE CAUSE OF AUTOSOMAL DOMINANT AND RECESSIVE OSTEOCHONDROPLASIS. (1) THE AUTOSOMAL DOMINANT FORM OF STICKLER SYNDROME (SS) IS CHARACTERIZED BY MILD SPONDYLOEPIPHYSEAL DYSPLASIA, OSTEOARTHRITIS, AND SENSORINEURAL HEARING LOSS. (2) THE AUTOSOMAL RECESSIVE DISORDER IS SIMILAR BUT MORE SEVERE.

-!- DISEASE: DEFECTS IN COL1A2 ARE THE CAUSE OF AUTOSOMAL DOMINANT NONSYNDROMIC SENSORINEURAL DEAFNESS TYPE 13 (DNA13). AFFECTED INDIVIDUALS EXPERIENCE PROGRESSIVE HEARING LOSS BEGINNING IN THE SECOND TO FOURTH DECADES, EVENTUALLY MAKING USE OF AMPLIFICATION MANDATORY.

-!- SIMILARITY: BELONGS TO THE FIBRILLAR CLASS OF COLLAGENS.

-!- DATABASE: NAME="Hereditary hearing loss homepage"; NOTE="Gene page"; WWW="http://www.via.ac.be/analab/fhh/fhhgenes.html".

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/>).

FT	PROPEP	1501	1736	CARBOXYL-TERMINAL PROPEPTIDE.
FT	DOMAIN	23	255	NONHELICAL REGION.
FT	DOMAIN	487	1500	TRIPLE HELICAL REGION.
FT	DOMAIN	1501	1736	NONHELICAL REGION.
FT	DOMAIN	10	18	POLY-LEU.
FT	CARBOHYD	298	301	POLY-GLU.
FT	CARBONYD	1604	1604	N-LINKED (GlcNAc . . .) (POTENTIAL).
FT	VARSPLIC	267	292	MISSING (IN ISOFORMS WITHOUT EXON 5).
FT	VARSPLIC	293	313	MISSING (IN ISOFORMS WITHOUT EXON 7).
FT	VARSPLIC	314	373	MISSING (IN ISOFORMS WITHOUT EXON 8).
FT	VARIANT	661	661	G -> R IN AUTOSOMAL RECESSIVE OSTEochondroDYSPLASIA.
<i>/FTid=VAR_0119073.</i>				
FT	VARIANT	808	808	G -> E (IN DNA13).
FT	CONFFLICT	97	97	<i>/FTid=VAR_010655.</i>
FT	CONFFLICT	531	531	R -> C (IN DNA13).
FT	CONFFLICT	542	542	<i>/FTid=VAR_010656.</i>
FT	CONFFLICT	548	549	A -> P (IN REF. 3).
FT	CONFFLICT	578	579	MP -> TL (IN REF. 3).
FT	CONFFLICT	835	835	AQ -> PR (IN REF. 3).
FT	CONFFLICT	835	835	S -> P (IN REF. 3 AND 5).
FT	CONFFLICT	97	97	Q -> R (IN REF. 3 AND 5).
FT	CONFFLICT	531	531	SL -> PP (IN REF. 2).
FT	CONFFLICT	542	542	<i>/FTid=VAR_010656.</i>
FT	CONFFLICT	548	549	TGPR -> HGST (IN REF. 4).
FT	CONFFLICT	842	842	QGP -> SG5 (IN REF. 4).
FT	CONFFLICT	894	894	P -> L (IN REF. 3).
FT	CONFFLICT	704	705	NO -> RP (IN REF. 3).
FT	CONFFLICT	704	704	R -> Q (IN REF. 3).
FT	CONFFLICT	726	726	D -> N (IN REF. 3).
FT	CONFFLICT	726	726	<i>/FTid=VAR_010655.</i>
FT	CONFFLICT	843	846	TGPR -> HGST (IN REF. 4).
FT	CONFFLICT	882	884	QGP -> SG5 (IN REF. 4).
FT	CONFFLICT	894	894	P -> L (IN REF. 1).
FT	CONFFLICT	1031	1032	PP -> RP (IN REF. 1 AND 4).
FT	CONFFLICT	1091	1091	D -> V (IN REF. 4).
FT	CONFFLICT	1124	1124	A -> R (IN REF. 4).
FT	CONFFLICT	1127	1133	EGARGP -> GAGGGT (IN REF. 4).
FT	CONFFLICT	1253	1253	A -> P (IN REF. 2).
FT	CONFFLICT	1257	1257	Q -> T (IN REF. 2).
FT	CONFFLICT	1552	1552	E -> R (IN REF. 4).
FT	CONFFLICT	1736	1736	RP (IN REF. 1 AND 4).
FT	CONFFLICT	1736	1736	MW: 722291568154B5F CRC64;
SQ SEQUENCE				
Query Match 59.8%; Score 55; DB 1; Length 1736;				
Best Local Similarity 66.7%; Pred. 5; Matches 10; Conservative 66%; Mismatches 0; Indels 5; Gaps 0; Gaps 0;				
Matches 10; Conservative 66%; Mismatches 0; Indels 5; Gaps 0; Gaps 0;				
RESULTS 21				
YK61_CAEEL	STANDARD:			
ID YK61_CAEEL	PRT:	305 AA.		
AC P34340;				
DT 01-FEB-1994 (Rel. 28, Last sequence update)				
DT 01-FEB-1994 (Rel. 28, Last sequence update)				
DT 01-MAR-2002 (Rel. 41, Last annotation update)				
DE Putative cuticle collagen C29E4.1.				
GN C29E4.1.				
OS Caenorhabditis elegans.				
OC Rhabditidae; Peioderinae; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;				
OC Rhabditida; Peioderinae; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;				
OX NCBITaxID:6239;				
RN [1]				
RP SEQUENCE FROM N A.				
RC STRAIN=BRIStOL N2;				
RA MEDLINE=94150718; PubMed=7906398;				
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M., Coulson A., Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A., Craxton M., Dear S., Du Z., Durbin R., Favell A., Fraser A., Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnstone L., Jones M., Kershaw J., Kirsten J., Lainster N., Latreille P., Lightning J., Lloyd C., Mortimore B., O'Gallaghan M., Parsons J., Percy C., Riffen L., Roopra A., Saunders D., Shownkeen R., Sims M., Smalton N., Smith A., Smith M., Sonnhammer E., Staden R.,				

RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,  
 RA Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,  
 RA Wohldman P.;  
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
 elegans";  
 RL Nature 368:37-38(1994).  
 -1- FUNCTION: NEURODE CUTICLES ARE COMPOSED LARGELY OF COLLAGEN-LIKE  
 PROTEINS. THE CUTICLE FUNCTIONS BOTH AS AN EXOSKELETON AND AS A  
 BARRIER TO PROTECT THE WORM FROM ITS ENVIRONMENT (BY SIMILARITY).  
 -1- SUBUNIT: COLLAGEN POLYPEPTIDE CHAINS ARE COMPLEXED WITHIN THE  
 CUTICLE BY DISULFIDE BONDS AND OTHER TYPES OF COVALENT CROSS-  
 LINKS (BY SIMILARITY).  
 CC COLLAGENS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 or send an email to license@isb-sib.ch).

CC EMBL; L23651; AAA27958; 1; -;  
 DR PIR; S44767; S44767;  
 DR WormPep; C29E4\_1; CE00083;  
 DR InterPro; IPR002480; Col\_cuticle\_N;  
 DR Pfam; PF01391; Collagen; 2;  
 DR Pfam; PF01484; Col\_cuticle\_N; 1;  
 KW Hypothetical protein; Cuticle; Connective tissue; Repeat;  
 KW Multigene family; Collagen.  
 FT DOMAIN 96 125 TRIPLE-HELICAL REGION.  
 FT DOMAIN 142 204 TRIPLE-HELICAL REGION.  
 FT DOMAIN 208 252 TRIPLE-HELICAL REGION.  
 FT DOMAIN 256 270 TRIPLE-HELICAL REGION.  
 SO SEQUENCE 305 AA; 30623 MW; BA4CCED5501E27D CRC64;

Query Match 58.7%; Score 54; DB 1; Length 305;  
 Best Local Similarity 66.7%; Pred No. 1.4;  
 Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
 Db 140 PAGGAGPGQQGPVG 154

RESULT 22  
 CA13\_BOVIN  
 ID CA13\_BOVIN STANDARD: PRT: 1049 AA.  
 AC P04258;  
 DT 20-MAR-1987 (Rel. 04, created)  
 DT 20-MAR-1987 (Rel. 04, last sequence update)  
 DT 16-OCT-2001 (Rel. 40, last annotation update)  
 DE Collagen alpha 1(III) chain.  
 GN COL3AI.  
 OS Bos taurus (Bovine).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
 OC Bovidae; Bovinae; Bos.  
 NCBI\_TAXID=9913;  
 RN [1]  
 SEQUENCE OF 1-242.

RX MEDLINE=80026026; PubMed=488906;  
 RX Fietzek P.P., Altmann H., Rauterberg J., Henkel W., Wachter E.,  
 RA Kuhn K.;  
 RT "The covalent structure of calf skin type III collagen. I-The amino acid sequence of the amino terminal region of the alpha 1(III) chain  
 RT (positions 1-222);"  
 RL Hoppe-Seyler's Z. Physiol. Chem. 360:809-820(1979).  
 RN [2]  
 SEQUENCE OF 243-422.

RX MEDLINE=80026027; PubMed=488907;

CC Dewes H., Fietzek P.P., Kuhn K.;  
 RT "The covalent structure of calf skin type III collagen. II. The amino acid sequence of the cyanogen bromide peptide alpha 1(III)CB1,8,10,2 (positions 223-402);";  
 RT Hoppe-Seyler's Z. Physiol. Chem. 360:821-832(1979).  
 RL RN [3]  
 RP SEQUENCE OF 423-571.  
 RX MEDLINE=80026028; PubMed=488908;  
 RA Bentz H., Fietzek P.P., Kuhn K.;  
 RT "The covalent structure of calf skin type III collagen. III. The amino acid sequence of the cyanogen bromide peptide alpha 1(III)CB4 (positions 403-551)";  
 RL Hoppe-Seyler's Z. Physiol. Chem. 360:833-840(1979).  
 RN [4]  
 RP SEQUENCE OF 572-808.  
 RX MEDLINE=80026029; PubMed=488909;  
 RA Lang H., Gianville R.W., Fietzek P.P., Kuhn K.;  
 RT "The covalent structure of calf skin type III collagen. IV. The amino acid sequence of the cyanogen bromide peptide alpha 1(III)CB5 (positions 555-788).";  
 RT Hoppe-Seyler's Z. Physiol. Chem. 360:841-850(1979).  
 RL RN [5]  
 RP SEQUENCE OF 809-947.  
 RX MEDLINE=80026030; PubMed=488910;  
 RA Dewes H., Fietzek P.P., Kuhn K.;  
 RT "The covalent structure of calf skin type III collagen. V. The amino acid sequence of the cyanogen bromide peptide alpha 1(III)CB9A (position 789-927).";  
 RT Hoppe-Seyler's Z. Physiol. Chem. 360:851-860(1979).  
 RL RN [6]  
 RP SEQUENCE OF 948-1049.  
 RX MEDLINE=80026031; PubMed=488911;  
 RA Altmann H., Fietzek P.P., Gianville R.W., Kuhn K.;  
 RT "The covalent structure of calf skin type III collagen. VI. The amino acid sequence of the carboxy-terminal cyanogen bromide peptide alpha 1(III)CB9B (positions 928-1028).";  
 RT Hoppe-Seyler's Z. Physiol. Chem. 360:861-868(1979).  
 -1- FUNCTION: COLLAGEN TYPE III OCCURS IN MOST SOFT CONNECTIVE TISSUES  
 -1- ALONG WITH TYPE I COLLAGEN.  
 CC -1- SUBUNIT: TRIMERS OF IDENTICAL ALPHA 1(III) CHAINS. THE CHAINS ARE  
 CC LINKED TO EACH OTHER BY INTERCHAIN DISULFIDE BONDS. TRIMERS ARE  
 CC ALSO CROSS-LINKED VIA HYDROXYLYSINES.  
 CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPÉPIPE REPEATING  
 UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
 DR PIR; A02862; CGBOTS.  
 DR InterPro; IPR00087; Collagen.  
 DR InterPro; IPR01007; VWC.  
 DR Pfam; PF01391; Collagen; 17;  
 DR PROSITE; PS01208; VWF\_C; PARTIAL.  
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
 KW Glycoprotein; Collagen.  
 FT DOMAIN 1 14 NONHLCAL REGION (N-TERMINAL).  
 FT DOMAIN 15 1040 TRIPLE-HELICAL REGION.  
 FT DOMAIN 1041 1049 NONHLCAL REGION (C-TERMINAL).  
 FT MOD\_RES 95 95 HYDROXYLATION.  
 FT MOD\_RES 107 107 HYDROXYLATION.  
 FT MOD\_RES 119 119 HYDROXYLATION.  
 FT MOD\_RES 938 938 HYDROXYLATION.  
 FT MOD\_RES 950 950 HYDROXYLATION.  
 FT CARBOHYD 107 107 O-LINKED (GAL, . , ).  
 FT CARBOHYD 950 950 O-LINKED (GAL, . , ).  
 FT DISULFDID 1040 1040 INTERCHAIN.  
 FT DISULFDID 1041 1041 INTERCHAIN.  
 SQ SEQUENCE 1049 AA; 93651 MW; 8BECC33DC66FC9A3 CRC64;

Query Match 58.7%; Score 54; DB 1; Length 1049;  
 Best Local Similarity 66.7%; Pred No. 4.3;  
 Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 Db 985 PVGSGPCKDGASG 999

RESULT 23  
 CA21\_RANCA STANDARD; PRT; 1355 AA.  
 ID CA21\_RANCA STANDARD; PRT; 1355 AA.  
 AC 042350; DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DE Collagen alpha 2(I) chain precursor.  
 GN COLIA2.  
 OS Rana catesbeiana (Bull frog).  
 OC Amphibia; Batrachia; Chordata; Craniata; Vertebrate; Euteleostomi;  
 NCBI\_TaxID=8400;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE="tail".  
 RX MEDLINE=9417499; PubMed=9272872;  
 RA Asahina K.; Oofusa K.; Obara M.; Yoshizato K.;  
 RT "Cloning and characterization of the full length cDNA encoding alpha 2  
 type I collagen of bullfrog Rana catesbeiana.";  
 RL Gene 194:283-289(1997).  
 CC -!- FUNCTION: TYPE I COLLAGEN IS A MEMBER OF GROUP I COLLAGEN  
 (FIBRILLAR FORMING COLLAGEN).  
 CC -!- SUBUNIT: TRIMERS OF ONE ALPHA 2(I) AND TWO ALPHA 1(I) CHAINS.  
 CC -!- TISSUE SPECIFICITY: FORMS THE FIBRILS OF TENDON, LIGAMENTS AND  
 BONES. IN BONES THE FIBRILS ARE MINERALIZED WITH CALCIUM  
 HYDROXYAPATITE.  
 CC -!- PROLINE AT THE THIRD POSITION OF THE TRIPETIDE REPEATING  
 UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>  
 or send an email to license@isb-sib.ch).  
 CC  
 CC  
 DR EMBL; D88764; BAA22380.1. --.  
 DR InterPro; IPR000885; Fib\_collagen\_C.  
 DR Pfam; PF01410; COLFI\_1.  
 DR ProDom; P000278; Fib\_collagen\_C; 1.  
 DR SMART; SM00038; COLFI\_1.  
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;  
 Glycoprotein; Collagen; Signal.  
 FT SIGNAL 1 22 POTENTIAL.  
 FT PROPEP 23 82 AMINO-TERMINAL PROPEPTIDE (POTENTIAL).  
 FT CHAIN 83 1093 COLLAGEN ALPHA 2(I) CHAIN.  
 FT PROPEP 1094 1355 CARBOXY-TERMINAL PROPEPTIDE  
 (BY SIMILARITY).  
 FT CARBOHYD 1205 1206 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1256 1256 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 1355 AA; 127643 MW; CB793AD56F11D2A CRC64;

Query Match 58.7%; Score 54; DB 1; Length 1355;  
 Best Local Similarity 60.0%; Pred. No. 5.5%;  
 Matches 9; Conservative 3; Mismatches 3; Indels 0; Gaps 0;  
 Qy 1 PAGPMQPNKGDKGV 15  
 Qy | :| || || ||:| |  
 Db 1046 PSGPSGGKGRSG 1060

RESULT 24  
 CA13\_HUMAN STANDARD; PRT; 1466 AA.  
 ID CA13\_HUMAN STANDARD; PRT; 1466 AA.  
 AC P02467; Q5112; DT 21-JUL-1986 (Rel. 01, Created)  
 DT 01-JAN-1990 (Rel. 13, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Collagen alpha 1(III) chain precursor.  
 GN COL3AI.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi;  
 OC Mammalia; Buteraria; Primates; Catarhini; Hominidae; Homo.  
 OC NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Skin fibroblast.  
 RX MEDLINE=89350838; PubMed=2764886;  
 RA Ala Kokko L.; Kontusari S.; Baldwin C.T.; Kuivaniemi H.;  
 Prockop D.J.;  
 RA "Structure of cDNA clones coding for the entire prepro alpha 1 (III)  
 chain of human type III procollagen. Differences in protein structure  
 from type I procollagen and conservation of codon preferences.";  
 RL Biochem. J. 260:509-516(1989).  
 RN [2]  
 RP SEQUENCE OF 149-1225 FROM N.A.  
 RX MEDLINE=89386015; PubMed=2780304;  
 RA Janecka R.A.; Ramirez F.;  
 RT "Nucleotide and amino acid sequences of the entire human alpha 1  
 (III) collagen";  
 RL Nucleic Acids Res. 17:6742-6742(1989).  
 RN [3]  
 RP SEQUENCE OF 168-398.  
 RX MEDLINE=77134724; PubMed=557335;  
 RA Seyer J.M.; Kang A.H.;  
 RT "Covalent structure of collagen: amino acid sequence of cyanogen  
 bromide peptides from the amino-terminal segment of type III collagen  
 of human liver";  
 RL Biochemistry 16:1158-1164(1977).  
 RN [4]  
 RP SEQUENCES.  
 RA Seyer J.M.;  
 RT Submitted (DEC-1977) to the PIR data bank.  
 RL Biochemistry 17:3404-3411(1978).  
 RN [5]  
 RP SEQUENCE OF 399-727.  
 RX MEDLINE=7900343; PubMed=687591;  
 RA Seyer J.M.; Kang A.H.;  
 RT "Covalent structure of collagen: amino acid sequence of five  
 consecutive CNBr peptides from type III collagen of human liver.>";  
 RL Biochemistry 17:2337-2337(1978).  
 RN [6]  
 RP SEQUENCE OF 728-964.  
 RX MEDLINE=80198282; PubMed=6246925;  
 RA Seyer J.M.; Mainardi C.; Kang A.H.;  
 RT "Covalent structure of collagen: amino acid sequence of alpha 1  
 (III)-CB5 from type III collagen of human liver.";  
 RL Biochemistry 19:1583-1589(1980).  
 RN [7]  
 RP SEQUENCE OF 950-1466 FROM N.A.  
 RX MEDLINE=8818982; PubMed=3357782;  
 RA Mankos B.S.; Dalglish R.;  
 RT "Human pro alpha 1(III) collagen: cDNA sequence for the 3' end.";  
 RL Nucleic Acids Res. 16:2337-2337(1988).  
 RN [8]  
 RP REVISION TO 1184.  
 RX MEDLINE=89098346; PubMed=3211760;  
 RA Molneux K.; Dalglish R.;  
 RT "Human type III collagen 'variant' is a cDNA cloning artefact.";  
 RL Nucleic Acids Res. 16:11833-11833(1988).  
 RN [9]  
 RP SEQUENCE OF 1065-1466 FROM N.A.  
 RX MEDLINE=89098794; PubMed=6096827;  
 RA Loidl H.R.; Brinker J.M.; May M.; Pihlajaniemi T.; Morrow S.;  
 RA Rosenblom J.; Myers J.C.;  
 RT "Molecular cloning and carboxyl-propeptide analysis of human type III  
 procollagen. ";  
 RL Nucleic Acids Res. 12:9383-9394(1984).  
 RN [10]  
 RP SEQUENCE OF 965-1200.  
 RX MEDLINE=81208139; PubMed=7016180;  
 RA Seyer J.M.; Kang A.H.;

- RT (III)-CB9 from type III collagen: amino acid sequence of alpha  
RL Blochimetry 20:2621-2627(1981).

RN [11] SEQUENCE OF 1176-1466 FROM N.A.  
RP MEDLINE=85157600; PubMed=2579949;  
RA Chu M.-L., Weil D., de Wet W.J., Bernard M.P., Sippola M., Ramirez F.;  
RT "Isolation of cDNA and genomic clones encoding human proalpha 1  
gene.";  
RT (III) collagen. Partial characterization of the 3' end region of the  
RT gene.";  
RL J. Biol. Chem. 260:4357-4363(1985).  
RN [12] SEQUENCE OF 1151-1200 FROM N.A.  
RP MEDLINE=86187804; PubMed=3154462;  
RA Miszkulin M., Daigle R., Klrove-Beckerman B., Rennard S.I.,  
RA Tolstoshev P., Brantly M., Crystal R.G.;  
RT Human type III collagen gene expression is coordinately modulated  
with the type I collagen genes during fibroblast growth.";  
RL Biochemistry 25:1408-1413(1986).  
RN [13] SEQUENCE OF 1-170 FROM N.A.  
RP TISSUE-Placenta;  
RX MEDLINE=88303360; PubMed=3405773;  
RA Toman D., Ricca G., de Crombrugghe B.;  
RT "Nucleotide sequence of a cDNA coding for the amino-terminal region  
of human prepro alpha 1(III) collagen.";  
RT Nucleic Acids Res. 16:7201-7201(1988).  
RN [14] SEQUENCE OF 1-176 FROM N.A.  
RP MEDLINE=89378752; PubMed=277083;  
RA Benson-Chanda V., Su M.W., Weil D., Chu M.-L., Ramirez F.;  
RT "Cloning and analysis of the 5' portion of the human type-III  
procollagen gene (COL3A1).";  
RL Gene 78:255-265(1989).  
RN [15] REVIEW ON VARIANTS.  
RP MEDLINE=97255959; PubMed=9101290;  
RA Kuivaniemi H., Tromp G., Prockop D.J.;  
RT "Mutations in fibrillar collagens (types I, II, III, and XI), fibrill-  
associated collagen (type IX), and network-forming collagen (type X)  
cause a spectrum of diseases of bone, cartilage, and blood vessels.";  
RL Hum. Mutat. 9:300-315(1997).  
RN [16] VARIANT AORTIC ANEURYSM ARG-303, AND VARIANT THR-668.  
RP MEDLINE=9329388; PubMed=8514866;  
RA Tromp G., Wu Y., Prockop D.J., Madhatheri S.L., Kleinert C.,  
RA Barley J.J., Zhuang J., Noerregaard O., Darling R.C., Abbott W.M.,  
Cole C.W., Jaakkola P., Rynnaenen M., Pearce W.H., Yao J.S.T.,  
RA Majamaa K., Smulmans S.N., Gatalica Z., Ferrell R.E., Jimenez S.A.,  
RA Jackson C.E., Michels V.V., Kaye M., Kuivaniemi H.;  
RT "Sequencing of cDNA from 50 unrelated patients reveals that mutations  
in the triple-helical domain of type III procollagen are an  
infrequent cause of aortic aneurysms.";  
RL J. Clin. Invest. 91:2539-2545(1993).  
RN [17] VARIANT THR-698.  
RP MEDLINE=91045136; PubMed=2235526;  
RA Wu Y., Ganalyn A., Prockop D.J.;  
RT "G to A polymorphism in exon 31 of the COL3A1 gene.";  
RL Nucleic Acids Res. 18:6180-6180(1990).  
RN [18] VARIANT AORTIC ANEURYSM ARG-786.  
RP MEDLINE=91056145; PubMed=2243125;  
RA Kontusaari S., Tromp G., Kuivaniemi H., Romanic A.M., Prockop D.J.;  
RT "A mutation in the gene for type III procollagen (COL3A1) in a family  
with aortic aneurysms.";  
RL J. Clin. Invest. 86:1465-1473(1990).  
RN [19] VARIANT EDS-IV ARG-828.  
RP MEDLINE=94016385; PubMed=8411057;  
RA Richards A.J., Narcisi P., Lloyd J.C., Ferguson C., Pope F.M.;  
RT "The substitution of glycine 661 by arginine in type III collagen  
[1]"

RT produces mutant molecules with different thermal stabilities and  
RT causes Ehlers-Danlos syndrome type IV.";  
RL ;  
RN [20] J. Med. Genet. 30:690-693(1993).  
RP VARIANT EDS-IV SER-957.  
RX MEDLINE=88109135; PubMed=2492273;  
RA Tromp G., Kuivaniemi H., Shikata H., Prockop D.J.;  
RT "A single base mutation that substitutes serine for glycine 790 of  
the alpha 1 (III) chain of type III procollagen exposes an arginine  
and causes Ehlers-Danlos syndrome type IV.";  
RL J. Biol. Chem. 264:1349-1352(1989).  
RN [21] J. Biol. Chem. 264:1349-1352(1989).  
RN [22] VARIANT EDS-IV VAL-960.  
RP MEDLINE=9316511; PubMed=7749417;  
RA Richards A.J., Ward P.N., Narcisi P., Nicholls A.C., Lloyd J.C.,  
RA Pope F.M.;  
RT "A single base mutation in the gene for type III collagen (COL3A1)  
converts glycine 847 to glutamic acid in a family with Ehlers-Danlos  
syndrome type IV. An unaffected family member is mosaic for the  
mutation.";  
RT Hum. Genet. 89:414-418(1992).  
RN [23] VARIANT EDS-IV ASP-1050.  
RP MEDLINE=91037070; PubMed=2808425;  
RA Tromp G., Kuivaniemi H., Stolle C.A., Pope F.M., Prockop D.J.;  
RT "Single base mutation in the type III procollagen gene that converts  
the codon for glycine 883 to aspartate in a mild variant of  
RT Ehlers-Danlos syndrome IV";  
RL J. Biol. Chem. 264:19313-19317(1989).  
RN [24] VARIANT EDS-IV VAL-1077.  
RP MEDLINE=9337480; PubMed=1895316;  
RA Richards A.J., Lloyd J.C., Ward P.N., de Paepe A., Narcisi P.,  
RA Pope F.M.;  
RT "Characterisation of a glycine to valine substitution at amino acid  
RT position 910 of the triple helical region of type III collagen in a  
patient with Ehlers-Danlos syndrome type IV.";  
RL J. Med. Genet. 28:458-463(1991).  
RN [25] VARIANT EDS-IV GLU-1173.  
RP MEDLINE=9022543; PubMed=1357323;  
RA Johnson P.H., Richards A.J., Pope F.M., Hopkinson D.A.;  
RT Quary Match 58.7%; Score 54; DB 1; Length 1466;  
RT Best Local Similarity 66.7%; Pred. No. 5,9; Mismatches 5; Indels 0; Gaps 0;  
RT Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
RN QY 1 PAGPQNGPQNGDKVG 15  
Db 1141 PVGPSSGGPKDGKVG 1155

RP	SEQUENCE FROM N.A. (SHORT ISOFORM).	DR	EMBL: U34609; AAC52901.1; JOINED.
RC	STRAIN=BALB/C; TISSUE=LIVER;	DR	EMBL: U34610; AAC52901.1; JOINED.
RX	MEDLINE=94245707; PubMed=8188673;	DR	EMBL: U34611; AAC52901.1; JOINED.
RA	Rehn M.V.; Hintikka E.; Pihlajaniemi T.;	DR	EMBL: U34612; AAC52901.1; JOINED.
RT	"Primary structure of the alpha 1 chain of mouse type XVIII collagen, partial structure of the corresponding gene, and comparison of the alpha 1(XVIII) chain with its homologue, the alpha 1(XV) collagen chain."	DR	EMBL: U34613; AAC52901.1; JOINED.
RT	J. Biol. Chem. 269:13929-13935(1994).	DR	EMBL: U03718; AAC52901.1; JOINED.
RT	[2]	DR	EMBL: U03715; AAC52902.1; JOINED.
RA	SEQUENCE FROM N.A. (SHORT AND LONG ISOFORMS).	DR	EMBL: U34608; AAC52902.1; JOINED.
RA	Rehn M.; Hintikka E.; Pihlajaniemi T.;	DR	EMBL: U34609; AAC52902.1; JOINED.
RT	"Characterization of the mouse gene for the alpha-1 chain of type XVIII collagen (COL18A1) reveals that the three variant N-terminal polypeptide forms are transcribed from two widely separated promoters."	DR	EMBL: U34610; AAC52902.1; JOINED.
RT	Submitted (NOV-1993) to the EMBL/genBank/DDJB databases.	DR	EMBL: U34611; AAC52902.1; JOINED.
RN	[3]	DR	EMBL: U34612; AAC52902.1; JOINED.
RN	SEQUENCE OF 213-1140 FROM N.A. (SHORT ISOFORM).	DR	EMBL: U34613; AAC52902.1; JOINED.
RX	MEDLINE=94240112; PubMed=8188894;	DR	EMBL: U34614; AAC52902.1; JOINED.
RA	Rehn M.V.; Pihlajaniemi T.;	DR	EMBL: U34615; AAC52902.1; JOINED.
RT	"Alpha 1(XVIII), a collagen chain with frequent interruptions in the collagenous sequence, a distinct tissue distribution, and homology with type XV collagen."	DR	EMBL: U03716; AAC52902.1; JOINED.
RT	proc. Natl. Acad. Sci. U.S.A. 91:4234-4238(1994).	DR	EMBL: U03717; AAC52902.1; JOINED.
RN	[4]	DR	EMBL: U11636; AAC52178.1; -.
RP	SEQUENCE OF 240-1527 FROM N.A.	DR	PDB: 1KOE; 16_FEB-99.
RC	TISSUE=Liver;	DR	MGI: 88451; Col18a1.
RX	MEDLINE=94240111; PubMed=8188893;	DR	InterPro; IPR000087; Collagen.
RA	On S.P., Kamakata Y., Muragaki Y., Timmons S., Ooshima A., Olsen B.R.; "Isolation and sequencing of cDNAs for proteins with multiple domains of GLY-Xaa-Yaa repeats identify a distinct family of collagenous proteins"; Proc. Natl. Acad. Sci. U.S.A. 91:4229-4233(1994).	DR	InterPro; IPR001791; Laminin_G.
RN	[5]	DR	InterPro; IPR003129; TSPN.
RP	CHARACTERIZATION OF ENDOSTATIN, AND PARTIAL SEQUENCE.	DR	Pfam; PF01391; Collagen; 8.
RX	MEDLINE=971160848; PubMed=9008168;	DR	SMART; SW00210; TSPN; 1.
RA	O'Reilly M.S., Boehm T., Shing Y., Fukai N., Vassios G., Lane W.S., Flynn E., Birkhead J.R., Olsen B.R., Folkman J.; "Endostatin: an endogenous inhibitor of angiogenesis and tumor growth"; Cell 88:277-285(1997).	DR	SM00282; LamG; 1.
RL	[6]	DR	SM00210; TSPN; 1.
RP	X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS) OF ENDOSTATIN.	KW	Extracellular matrix; Collagen; Glycoprotein; Signal; Alternative splicing; Cell adhesion; Connective tissue; Repeat; Hydroxylation; 3D-structure.
FT	SIGNAL	KW	POTENTIAL.
FT	CHAIN	FT	COLLAGEN ALPHA 1 (XVIII) CHAIN.
FT	CHAIN	FT	ENDOSTATIN.
FT	DOMAIN	FT	NONHELCICAL REGION 1 (NC1).
FT	DOMAIN	FT	TRIPLE-HELICAL REGION 1 (COL1).
FT	DOMAIN	FT	NONHELCICAL REGION 2 (NC2).
FT	DOMAIN	FT	TRIPLE-HELICAL REGION 2 (COL2).
FT	DOMAIN	FT	NONHELCICAL REGION 3 (NC3).
FT	DOMAIN	FT	TRIPLE-HELICAL REGION 3 (COL3).
FT	DOMAIN	FT	NONHELCICAL REGION 4 (NC4).
FT	DOMAIN	FT	TRIPLE-HELICAL REGION 4 (COL4).
FT	DOMAIN	FT	NONHELCICAL REGION 5 (NC5).
FT	DOMAIN	FT	TRIPLE-HELICAL REGION 5 (COL5).
FT	DOMAIN	FT	NONHELCICAL REGION 6 (NC6).
FT	DOMAIN	FT	TRIPLE-HELICAL REGION 6 (COL6).
FT	DOMAIN	FT	NONHELCICAL REGION 7 (NC7).
FT	DOMAIN	FT	TRIPLE-HELICAL REGION 7 (COL7).
FT	DOMAIN	FT	NONHELCICAL REGION 8 (NC8).
FT	DOMAIN	FT	TRIPLE-HELICAL REGION 8 (COL8).
FT	DOMAIN	FT	NONHELCICAL REGION 9 (NC9).
FT	DOMAIN	FT	TRIPLE-HELICAL REGION 9 (COL9).
FT	DOMAIN	FT	NONHELCICAL REGION 10 (NC10).
FT	DOMAIN	FT	TRIPLE-HELICAL REGION 10 (COL10).
FT	CARBOHYD	FT	NONHELCICAL REGION 11 (NC11).
FT	CARBONYD	FT	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBONYD	FT	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	DISULFID	FT	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	DISULFID	FT	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	SITE	FT	CELL ATTACHMENT SITE (POTENTIAL).
FT	VARSLIC	FT	MISSING (IN SHORT ISOFORM).
FT	VARSLIC	FT	AVPQQLPPFQSNLQPLGRSAPPF -> MAPRWHLDDVL
FT	VARSLIC	FT	TSVLLVVARVSWAE (IN SHORT ISOFORM).
FT	CONFICT	FT	P -> L (IN REF. 4).
FT	CONFICT	FT	P -> F (IN REF. 4).
FT	CONFICT	FT	A -> R (IN REF. 4).
FT	CONFICT	FT	R -> P (IN REF. 4).
FT	CONFICT	FT	P -> L (IN REF. 4).
FT	CONFICT	FT	L -> F (IN REF. 4).
FT	CONFICT	FT	L -> V (IN REF. 4).
SQ	SEQUENCE	FT	1527 AA; 156008 MW; 9645045AF140B513 CRC64;
		DR	58.7%; Score 54; DB 1; Length 1527;
Query Match			

	BEST LOCAL SIMILARITY 66.7%; PRED. NO. 6.1; MATCHES 10; CONSERVATIVE 0; MISMATCHES 5;	CC
QY	1 PAGPWGPCKDGKVG 15	CC
		CC
DB	580 PEPQGPGKDGPG 594	CC
RESULT 26		CC
CAT5_HUMAN	STANDARD;	CC
ID CAT5_HUMAN	PRT; 1838 AA.	CC
AC P0908:	01-FEB-1991 (Rel. 17, Created)	CC
DT 01-MAY-1992 (Rel. 22, Last sequence update)	CC	
DT 16-OCT-2001 (Rel. 40, Last annotation update)	CC	
DE Collagen alpha 1(V) chain precursor.	CC	
GN COL5A1.	CC	
OS Homo sapiens (Human).	CC	
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	CC	
OC NCBI_TAXID=9606;	CC	
RN [1]		CC
SEQUENCE FROM N.A., AND SEQUENCE OF 556-565.	CC	
RX MEDLINE=91302336; PubMed=2071595;	CC	
RA Takahara K., Seto Y., Okamoto N., Noda A., Yaoi Y., Kato I.;	CC	
RA "Complete primary structure of human collagen alpha 1 (V) chain.", J. Biol. Chem. 266:13124-13129(1991).	CC	
RP [2]		CC
RC TISSUE=Chorionamniotic membrane;	CC	
RX MEDLINE=89227189; PubMed=2496661;	CC	
RA Sever J.M., Kang A.H.;	CC	
RT "Covalent structure of collagen: amino acid sequence of three cysteine bromide-derived peptides from human alpha 1(V) collagen chain.", Arch. Biochem. Biophys. 271:120-129(1989).	CC	
RL RN [3]		CC
RP SEQUENCE OF 823-950, AND HEPARIN-BINDING.	CC	
RX MEDLINE=90366601; PubMed=2203476;	CC	
RA Yaoi Y., Hashimoto K., Koitabashi H., Takahara K., Ito M., Kato I.;	CC	
RT "Primary structure of the heparin-binding site of type V collagen.", Biochim. Biophys. Acta 1035:139-145(1990).	CC	
RL RN [4]		CC
RP SEQUENCE OF 556-571.	CC	
RC TISSUE=placenta;	CC	
RX MEDLINE=92239022; PubMed=1571108;	CC	
RA Mann K.;	CC	
RT "Isolation of the alpha 3-chain of human type V collagen and characterization by partial sequencing.", Biol. Chem. Hoppe-Seyler 373:69-75(1992).	CC	
RL RN [5]		CC
RP SEQUENCE OF 565-576; 756-772; 1012-1029; 1219-1232 AND 1465-1477.	CC	
RC TISSUE=chorionamniotic membrane;	CC	
RX MEDLINE=94237164; PubMed=8181482;	CC	
RA Moradi-Kamel M., Rousseau J.C., Kleman J.P., Champialaud M.F., Bontillon M.M., Bernillon J., Wallach J.M., van der Rest M.;	CC	
RT "Diversity in the processing events at the N-terminus of type-V collagen.", Eur. J. Biochem. 221:987-995(1994).	CC	
RL RN [6]		CC
RP VARIANT EDS1 SER-1639.	CC	
RX MEDLINE=97195540; PubMed=902913;	CC	
RA de Paepse A., Nuytinck L., Haussler I., Anton-Lamprecht I., Naeyaert J.-M.;	CC	
RT "Mutations in the COL5A1 gene are causal in the Ehlers-Danlos syndromes I and II.", Am. J. Hum. Genet. 60:547-554(1997).	CC	
RL CC -1- FUNCTION: TYPE V COLLAGEN IS A MEMBER OF GROUP I COLLAGEN (FIBRILLAR FORMING COLLAGEN). IT IS A MINOR CONNECTIVE TISSUE COMPONENT OF NEARLY UBIQUITOUS DISTRIBUTION. TYPE V COLLAGEN BINDS TO DNA, HEPARAN SULFATE, THROMBOSONDIN, HEPARIN, AND INSULIN.	CC	
CC -1- SUBUNIT: TRIMERS OF TWO ALPHA 1(V) AND ONE ALPHA 2(V) CHAINS IN	CC	
	MOST TISSUES AND TRIMERS OF ONE ALPHA 1(V), ONE ALPHA 2(V), AND ONE ALPHA 3(V) CHAINS IN PLACENTA.	CC
	-1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPETIDE REPEATING UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.	CC
	-1- PTM: 40% OF TYROSINES IN THE PRO-ALPHA 1(V) CHAIN ARE SULFATED.	CC
	-1- DISEASE: DEFECTS IN COL5A1 ARE A CAUSE OF OEHLERS-DANLOS SYNDROME, TYPE I (EDSI), A DISEASE CHARACTERIZED BY LOOSE-JOINTEDNESS AND FRAGILE, VELVETY, STRETCHABLE, BRUISEABLE SKIN THAT HEALS WITH PECULIAR 'CIGARETTE-PAPER' SCARS.	CC
	-1- SIMILARITY: HIGH, TO ALPHA 3(V) AND ALPHA 1(XI) CHAINS.	CC
	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a> or send an email to license@isb-sib.ch).	CC
	-----	CC
	DR DR0279; BAA14323.1; -.	CC
	DR P03978; S03978.	CC
	DR S11303; S11303.	CC
	DR S16024; S16024.	CC
	DR MTM; 120215; -.	CC
	DR MM; 130000; -.	CC
	DR MM; 130010; -.	CC
	DR InterPro; IPR000087; Collagen.	CC
	DR InterPro; IPR00885; Fib_collagen_C.	CC
	DR InterPro; IPR01191; Laminin_G.	CC
	DR InterPro; IPR03129; TSPPN.	CC
	DR Pfam; PF01410; COLFI; 1.	CC
	DR Pfam; PF01391; Collagen; 18.	CC
	DR PRODOM; PD002078; Fib_collagen_C; 1.	CC
	DR SMART; SM00038; COLFI; 1.	CC
	DR SMART; SM00822; LamG; 1.	CC
	DR SMART; SM00210; TSPPN; 1.	CC
	KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation; Collagen; Signal; Heparin-binding; Sulfation; Disease mutation.	CC
	FT SIGNAL 1 37.	CC
	FT CHAIN 38 1605	CC
	FT DOMAIN 38 443	CC
	FT DOMAIN 444 558	CC
	FT DOMAIN 559 1570	CC
	FT DOMAIN 1571 1605	CC
	FT PROPEP 1606 1838	CC
	FT MOD_RES 570 570	CC
	FT MOD_RES 576 576	CC
	FT MOD_RES 621 621	CC
	FT MOD_RES 627 627	CC
	FT MOD_RES 639 639	CC
	FT MOD_RES 642 642	CC
	FT MOD_RES 648 648	CC
	FT MOD_RES 654 654	CC
	FT MOD_RES 657 657	CC
	FT MOD_RES 675 675	CC
	FT MOD_RES 678 678	CC
	FT MOD_RES 680 680	CC
	FT MOD_RES 686 686	CC
	FT MOD_RES 690 690	CC
	FT MOD_RES 696 696	CC
	FT MOD_RES 705 705	CC
	FT MOD_RES 708 708	CC
	FT MOD_RES 717 717	CC
	FT MOD_RES 720 720	CC
	FT MOD_RES 726 726	CC
	FT MOD_RES 732 732	CC
	FT MOD_RES 744 744	CC
	FT MOD_RES 750 750	CC
	FT MOD_RES 756 756	CC
	FT MOD_RES 762 762	CC
	FT MOD_RES 765 765	CC
	FT MOD_RES 771 771	CC

FT	MOD_RES	774	774	HYDROXYLATION.
FT	MOD_RES	780	780	HYDROXYLATION.
FT	MOD_RES	789	789	HYDROXYLATION.
FT	MOD_RES	795	795	HYDROXYLATION.
FT	MOD_RES	804	804	HYDROXYLATION.
FT	MOD_RES	807	807	HYDROXYLATION.
FT	MOD_RES	810	810	HYDROXYLATION.
FT	MOD_RES	816	816	HYDROXYLATION.
FT	MOD_RES	819	819	HYDROXYLATION.
FT	MOD_RES	834	834	HYDROXYLATION.
FT	MOD_RES	846	846	HYDROXYLATION.
FT	MOD_RES	861	861	HYDROXYLATION.
FT	MOD_RES	864	864	HYDROXYLATION.
FT	MOD_RES	870	870	HYDROXYLATION.
FT	MOD_RES	873	873	HYDROXYLATION.
FT	MOD_RES	876	876	HYDROXYLATION.
FT	MOD_RES	882	882	HYDROXYLATION.
FT	MOD_RES	888	888	HYDROXYLATION.
FT	MOD_RES	891	891	HYDROXYLATION.
FT	MOD_RES	897	897	HYDROXYLATION.
FT	MOD_RES	903	903	HYDROXYLATION.
FT	MOD_RES	905	905	HYDROXYLATION.
FT	MOD_RES	930	930	HYDROXYLATION.
FT	MOD_RES	945	945	HYDROXYLATION.
FT	MOD_RES	1017	1017	HYDROXYLATION.
FT	MOD_RES	1020	1020	HYDROXYLATION.
FT	MOD_RES	1023	1023	HYDROXYLATION.
FT	MOD_RES	1029	1029	HYDROXYLATION.
FT	MOD_RES	1221	1221	HYDROXYLATION.
FT	MOD_RES	1224	1224	HYDROXYLATION.
FT	MOD_RES	1467	1457	HYDROXYLATION.
FT	MOD_RES	1470	1470	HYDROXYLATION.
FT	VARIANT	1639	1639	C -> S ( IN EDS1 ).
/FT	ID			/FT ID VAR_001808.
E	-> G ( IN REF. 2 ).			E -> G ( IN REF. 2 ).
P	-> L ( IN REF. 2 ).			P -> L ( IN REF. 2 ).
R	-> E ( IN REF. 2 ).			R -> E ( IN REF. 2 ).
E	-> Q ( IN REF. 2 ).			E -> Q ( IN REF. 2 ).
K	-> Q ( IN REF. 2 ).			K -> Q ( IN REF. 2 ).
L	-> P ( IN REF. 2 ).			L -> P ( IN REF. 2 ).
F	PGGPPGVT -> VTGEVGAP ( IN REF. 2 ).			F -> PG -> VTGEVGAP ( IN REF. 2 ).
T	CONFLICT	650	650	G -> Q ( IN REF. 2 ).
T	CONFLICT	663	663	P -> I ( IN REF. 2 ).
F	CONFLICT	668	668	E -> Q ( IN REF. 2 ).
F	CONFLICT	677	677	K -> Q ( IN REF. 2 ).
F	CONFLICT	684	684	L -> P ( IN REF. 2 ).
F	CONFLICT	699	699	PGGPPGVT -> VTGEVGAP ( IN REF. 2 ).
F	CONFLICT	727	727	G -> Q ( IN REF. 2 ).
F	CONFLICT	741	741	P -> I ( IN REF. 2 ).
F	CONFLICT	747	747	E -> Q ( IN REF. 2 ).
F	CONFLICT	753	753	K -> Q ( IN REF. 2 ).
F	CONFLICT	759	759	L -> P ( IN REF. 2 ).
F	CONFLICT	777	777	D -> N ( IN REF. 2 ).
F	CONFLICT	849	855	QO -> OK ( IN REF. 2 ).
F	CONFLICT	894	894	GGPNGD -> IGGPGR ( IN REF. 3 ).
SQ	SEQUENCE	1838 AA;	183616 MW;	N -> D ( IN REF. 3 ).
Query Match	58.7%	Score 54;	DB 1;	Length 1838;
Best Local Similarity	66.7%	Pred. No. 7.2;		
Matches	10;	Conservative	1;	Mismatches
Qy	1	PAGPWPPNGKCKGVG 15	Qy	1
Db	1145	PAGPVGGPGEDDKG 1159	Db	125
RESULT 27			RESULT 28	
CC36_CAEEL	STANDARD;	PRT;	CC06_CAEEL	STANDARD;
ID			ID	
AC			AC	
P34803;			P18831;	
01-FEB-1994	( Rel. 28, Last sequence update)		DT	01-NOV-1990 ( Rel. 16, Created )
DT	01-FEB-2001	( Rel. 40, Last annotation update )	DT	01-NOV-1990 ( Rel. 16, Last sequence update )
DE	Cuticule collagen 36.		DT	01-JUN-1994 ( Rel. 29, Last annotation update )
GN	COL-35 OR C2H5.5.		DE	Cuticule collagen 6.
OS	Caenorhabditis elegans.		GN	COL-6.
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;		OS	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC	Rhabditidae; Palaeoderrinae; Caenorhabditis.		OC	Rhabditidae; Palaeoderrinae; Caenorhabditis.
OX	NCBI_TaxID=6239;		OX	NCBI_TaxID=6239;
	[1]		RP	SEQUENCE FROM N.A.
			RP	STRAIN=BRISTOL N2;
			RX	MEDLINE=94131298; PubMed=8299960;
			RA	Levy A.D.; Kramer J.M.;
			RT	"Identification, sequence and expression patterns of the
			RT	Caenorhabditis elegans, col-36 and col-40 collagen-encoding genes. ";
			RL	Gene 137:281-285(1993).
			RL	(2)
			RP	SEQUENCE FROM N.A.
			RP	STRAIN=BRISTOL N2;
			RC	
			CC	Submitted (JUL-1995) to the EMBL/GenBank/DDBJ databases.
			CC	-1- FUNCTION: NEURITO CUTICLES ARE COMPOSED LARGELY OF COLLAGEN-LIKE
			CC	PROTEINS. THE CUTICLE FUNCTIONS BOTH AS AN EXOSKELETON AND AS A
			CC	BARRIER TO PROTECT THE WORM FROM ITS ENVIRONMENT.
			CC	-1- SUBUNIT: COLLAGEN POLYPEPTIDE CHAINS ARE COMPLEXED WITHIN THE
			CC	CUTICLE BY DISULFIDE BONDS AND OTHER TYPES OF COVALENT CROSS-
			CC	LINKS.
			CC	-1- SIMILARITY: TO OTHER COLLAGENS. STRONG, TO OTHER CUTICLE
			CC	COLLAGENS.
			CC	.
			CC	.
			CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
			CC	between the Swiss Institute of Bioinformatics and the EMBL outstation
			CC	the European Bioinformatics Institute. There are no restrictions on its
			CC	use by non-profit institutions as long as its content is in no way
			CC	modified and this statement is not removed. Usage by and for commercial
			CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce">http://www.isb-sib.ch/announce</a> or send an email to license@isb-sib.ch).
			CC	.
			DR	EMBL; U14635; ARK84452.1; -.
			DR	EMBL; C27HS.5; CEP6893.
			DR	InterPro; IPR002486; Col-cuticle_N.
			DR	InterPro; IPR000087; Collagen.
			DR	PFAM; PF01184; collagen_3.
			DR	PFAM; PF01184; collagen_N.
			KW	Cuticule; Connective tissue; Repete; Multigene family; Collagen.
			FT	DOMAIN 89 105 TRIPLE-HELICAL REGION.
			FT	DOMAIN 118 150 TRIPLE-HELICAL REGION.
			FT	DOMAIN 167 187 TRIPLE-HELICAL REGION.
			FT	DOMAIN 194 226 TRIPLE-HELICAL REGION.
			FT	DOMAIN 231 257 TRIPLE-HELICAL REGION.
			FT	DOMAIN 260 295 TRIPLE-HELICAL REGION.
			SQ	SEQUENCE 307 AA; 30126 MW; 9C7206CC18660FOB CRC64;
			Query Match	57.6%
			Best Local Similarity	66.7%
			Pred. No. 1.9;	Score 53;
			0;	DB 1;
			Mismatches	Length 307;
			5;	Indels 0;
			0;	Gaps 0;
			RESULT 28	
			CC06_CAEEL	STANDARD;
			ID	
			AC	
			P18831;	
			DT	01-NOV-1990 ( Rel. 16, Created )
			DT	01-NOV-1990 ( Rel. 16, Last sequence update )
			DT	01-JUN-1994 ( Rel. 29, Last annotation update )
			DE	Cuticule collagen 6.
			GN	COL-6.
			OS	Caenorhabditis elegans.
			OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
			OC	Rhabditidae; Palaeoderrinae; Caenorhabditis.
			OX	NCBI_TaxID=6239;
			RP	SEQUENCE FROM N.A.
			RP	STRAIN=BRISTOL N2;
			RX	MEDLINE=8926131; PubMed=2753356;
			RA	Cox G.N.; Fields C.; Kramer J.M.; Rosenzweig B.; Hirsh D.;
			RA	NCBI_TaxID=6239;

RT \*sequence comparisons of developmentally regulated collagen genes of  
 RT Caenorhabditis elegans.;  
 RL Gene 76:331-344 (1989).  
 CC -!- FUNCTION: NEMATODE CUTICLES ARE COMPOSED LARGELY OF COLLAGEN-LIKE  
 PROTEINS. THE CUTICLE FUNCTIONS BOTH AS AN EXOSKELETON AND AS A  
 CC BARRIER TO PROTECT THE WORM FROM ITS ENVIRONMENT.  
 CC -!- SUBUNIT: COLLAGEN POLYPEPTIDE CHAINS ARE COMPLEXED WITHIN THE  
 CUTICLE BY DISULFIDE BONDS AND OTHER TYPES OF COVALENT CROSS-  
 CC LINES.  
 CC -!- SIMILARITY: TO OTHER COLLAGENS. STRONG TO OTHER CUTICLE  
 CC COLLAGENS. COL-2 AND COL-6 BELONGS TO THE SAME GROUP OF COLLAGEN.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to license@isb-sib.ch).  
 CC  
 DR EMBL; M25477; AA227991.1; -.  
 DR InterPro; IPR002486; Col-cuticle\_N.  
 DR InterPro; IPR00087; Collagen.  
 DR Pfam; PF01391; Collagen; 2.  
 DR Pfam; PF01484; Col-cuticle\_N; 1.  
 KW Cuticle; Connective tissue; Repeat; Multigene family; Collagen.  
 FT DOMAIN 95 115 GY-RICH  
 FT DOMAIN 142 171 TRIPLE-HELICAL REGION.  
 FT DOMAIN 149 212 TRIPLE-HELICAL REGION.  
 FT DOMAIN 216 248 TRIPLE-HELICAL REGION.  
 FT DOMAIN 253 279 TRIPLE-HELICAL REGION.  
 FT DOMAIN 282 329 TRIPLE-HELICAL REGION.  
 SQ SEQUENCE 329 AA; 32561 MW; 52C295EF77E76418 CRC64;  
 Query Match 57.6%; Score 53; DB 1; Length 329;  
 Best Local Similarity 66.7%; Pred. No. 2;  
 Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 Qy | 1 PAGPWGPNGKDKKG 15  
 Db 146 PROPEGPQGNDGKG 160  
 RESULT 29  
 RSPD\_MOUSE STANDARD; PRT; 374 AA.  
 ID RSPD\_MOUSE  
 AC P0404;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Pulmonary surfactant-associated protein D precursor (SP-D) (PSP-D).  
 GN SPFD OR SFP4.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAINE-B6/CBA; TISSUE=Lung;  
 RX MEDLINE=96034460; PubMed=749852;  
 RA Motwani M., White R.A., Guo N., Dowler L.I., Tauber A.I., Sastry K.N.;  
 RT "Mouse surfactant protein-D, cDNA cloning, characterization, and gene  
 localization to chromosome 14.";  
 RL J. Immunol. 155:5671-5677 (1995).  
 CC -!- FUNCTION: CONTRIBUTES TO THE LUNG'S DEFENSE AGAINST INHALED  
 MICRORGANISMS. BINDS STRONGLY MALTOSE RESIDUES AND TO A LESSER  
 EXTENT OTHER ALPHA-GLUCOSYL MOIETIES. IT COULD PARTICIPATE IN THE  
 EXTRACELLULAR REORGANIZATION OR TURNOVER OF PULMONARY SURFACTANT.  
 CC -!- SUBUNIT: OLIGOMERIC COMPLEX OF 4 SET OF HOMOTRIMERS.  
 CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR.  
 CC -!- MISCELLANEOUS: PULMONARY SURFACTANT CONSISTS OF 90% LIPID AND 10%  
 PROTEIN. THERE ARE 4 SURFACTANT ASSOCIATED PROTEIN: 2 COLLAGENOUS,  
 CC  
 CC CARBOHYDRATE-BINDING GLYCOPROTEINS (SP-A AND SP-D) AND 2 SMALL  
 CC HYDROPHOBIC PROTEINS (SP-B AND SP-C).  
 CC -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to license@isb-sib.ch).  
 CC  
 DR EMBL; I40156; AAA92021.1; -.  
 DR HSSP; P35247; 1B08.  
 DR MGD; MGJ10515; Sftp4d  
 DR InterPro; IPR00087; Collagen.  
 DR InterPro; IPR001304; Lectin\_c.  
 DR Pfam; PF01391; Collagen; 3.  
 DR Pfam; PF00059; Lectin\_c; 1.  
 DR SMART; SM0034; CLECT; 1.  
 DR PROSITE; PS00615; C\_TYPE\_LECTIN\_1; 1.  
 DR PROSITE; PS00641; C\_TYPE\_LECTIN\_2; 1.  
 KW Glycoprotein; Calcium; Surface film; Gaseous exchange; Hydroxylation;  
 KW Signal; Lectin; Collagen; Repeat; Coiled coil.  
 FT SIGNAL 1 19 BY SIMILARITY.  
 CHAIN 20 374 PULMONARY SURFACTANT-ASSOCIATED PROTEIN  
 FT DOMAIN 45 221 D. COLLAGEN-LIKE.  
 FT DOMAIN 222 253 COILED COIL (POTENTIAL).  
 FT DOMAIN 278 374 C-TYPE LECTIN (SHORT FORM).  
 FT DISULFID 280 372 BY SIMILARITY.  
 FT DISULFID 350 364 BY SIMILARITY.  
 FT CARBOHYD 89 89 N-LINKED (GLCNAC- . . ) (POTENTIAL).  
 SQ SEQUENCE 374 AA; 37688 MW; FE034261263F43E4 CRC64;  
 Query Match 57.6%; Score 53; DB 1; Length 374;  
 Best Local Similarity 60.0%; Pred. No. 2.3;  
 Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
 Qy | 1 PAGPWGPNGKDGKVG 15  
 Db 115 PAGKEGPQGKQGNIG 129  
 RESULT 30  
 PSPD\_RAT STANDARD; PRT; 374 AA.  
 ID PSPD\_RAT  
 AC P35248;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Pulmonary surfactant-associated protein D precursor (SP-D) (PSP-D)  
 DE SPFD OR SFP4.  
 GN Rattus norvegicus (Rat).  
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 20-33.  
 RC TISSUE=Lung;  
 RX MEDLINE=92112913; PubMed=1370483;  
 RA Shinzui H., Fisher J.H., Popst P., Benson B., Lau K., Mason R.J.,  
 RA Voelker D.R.;  
 RT "Primary structure of rat pulmonary surfactant protein D. cDNA and  
 RT deduced amino acid sequence.";  
 RL J. Biol. Chem. 267:1855-1857 (1992).  
 RN [2]  
 RP SEQUENCE OF 73-95 AND 153-180.  
 RC TISSUE=Lung;  
 RX MEDLINE=9001186; PubMed=2675969;  
 RA Person A., Chang D., Rust K., Maxley M., Longmore W., Crouch E.;

RT Purification and biochemical characterization of CP4 (SP-D), a  
 RT collagenous surfactant associated protein.";  
 RL Biochemistry 28:6361-6367 (1989).  
 CC -!- FUNCTION: CONTRIBUTES TO THE LUNG'S DEFENSE AGAINST INHALED  
 CC MICROORGANISMS. BINDS STRONGLY MALTOSE RESIDUES AND TO A LESSER  
 CC EXTENT OTHER ALPHA-GLUCOSYL MOIETIES. IT COULD PARTICIPATE IN THE  
 CC EXTRACELLULAR REORGANIZATION OR TURNOVER OF PULMONARY SURFACTANT.  
 CC -!- SUBUNIT: OLIGOMERIC COMPLEX OF 4 SET OF HOMOTRIMERS.  
 CC -!- SUBCELLULAR LOCATION: Extracellular.  
 CC -!- MISCELLANEOUS: PULMONARY SURFACTANT CONSISTS OF 90% LIPTD AND 10%  
 CC PROTEIN. THERE ARE 4 SURFACTANT ASSOCIATED PROTEIN: 2 COLLAGENOUS,  
 CC CARBOHYDRATE-BINDING GLYCOPROTEINS (SP-A AND SP-D) AND 2 SMALL  
 CC HYDROPHOBIC PROTEINS (SP-B AND SP-C).  
 CC -!- SIMILARITY: CONTAINS A COLLAGENOUS DOMAIN.  
 CC -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to license@isb-sib.ch).  
 CC  
 DR EMBL; M81231; AAA42170.1; -.  
 DR PIR; A42046; A42046.  
 DR HSSP; P35247; 1B08.  
 DR InterPro; IPR000087; Collagen.  
 DR InterPro; IPR001304; lectin\_C.  
 DR Pfam; PF01391; Collagen; 3.  
 DR Pfam; PF00059; lectin\_c; 1.  
 DR SMART; SM0034; CLSGC; 1.  
 DR PROSITE; PS00615; C-TYPE\_LECTIN\_2; 1.  
 DR Glycoprotein; Calcium; Surface film; Gaseous exchange; Hydroxylation;  
 KW Signal; Lectin; Collagen; Repeat; Coiled coil.  
 FT SIGNAL 1 19  
 FT CHAIN 20 374  
 D. PULMONARY SURFACTANT-ASSOCIATED PROTEIN  
 FT DOMAIN 45 221  
 FT DOMAIN 222 253  
 FT DOMAIN 278 374  
 FT DISULFID 280 372  
 FT DISULFID 350 364  
 FT CARBOHYD 89 89  
 FT MOD\_RES 77 77  
 FT MOD\_RES 86 86  
 FT MOD\_RES 95 95  
 FT MOD\_RES 98 98  
 FT MOD\_RES 170 170  
 FT MOD\_RES 176 176  
 FT CONFLICT 89 89  
 FT CONFLICT 164 164  
 SQ SEQUENCE 374 AA; 37561 MW; DB2BB5E399DBA3C CRC64;  
 Query Match 57.6%; Score 53; DB 1; Length 374;  
 Best Local Similarity 60.0%; Pred. No. 2.3; Mismatches 4; Indels 0; Gaps 0;  
 Matches 9; Conservative 2;

Qy 1 PAGWGPNKGKVG 15  
 Db 115 PAGEGPGSKQGNITG 129

Search completed: November 1, 2002, 12:52:54  
 Job time : 8 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

On protein - protein search, using sw model  
Run on: November 1, 2002, 12:19:21 ; Search time 21.5 Seconds  
(without alignments)  
120.694 Million cell updates/sec

Title: US-09-529-691A-3  
Perfect score: 92  
Sequence: 1 PAGPWPNGKDGKV 15

Scoring table: BLOSUM2  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222  
Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

#### Database :

- 1: SP\_TREMBL\_19;\*
- 2: SP\_archaea;\*
- 3: SP\_bacteria;\*
- 4: SP\_fungi;\*
- 5: SP\_invertebrate;\*
- 6: SP\_mammal;\*
- 7: SP\_mhc;\*
- 8: SP\_organelle;\*
- 9: SP\_phage;\*
- 10: SP\_plant;\*
- 11: SP\_rabbit;\*
- 12: SP\_virus;\*
- 13: SP\_vertebrate;\*
- 14: SP\_unclassified;\*
- 15: SP\_rvirus;\*
- 16: SP\_bacterian;\*
- 17: SP\_archeap;\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query	Match Length	DB ID	Description	ALIGMENTS
1	69	75.0	460	5 Q9GF7	RESULT 1 ID: Q9GF7 AC: Q9GF7; PRIMINARY; PRF: 460 AA.	
2	66	71.7	291	5 Q19470	ID: Q9GF7; DT: 01-MAR-2001 (TREMBLrel. 16, Last sequence update)	
3	63	68.5	1366	4 Q15177	DT: 01-DEC-2001 (TREMBLrel. 19, Last annotation update)	
4	62	67.4	326	2 Q9AG66	DT: HYPOTHECTICAL 41.0 KDA PROTEIN.	
5	61	66.3	410	2 Q9F691	ID: C1BHT7.3; OS: Caenorhabditis elegans	
6	60	65.2	358	12 Q9J3U4	OC: Melazoa; Nematoidea; Rhabditida; Rhabditoidae;	
7	60	65.2	566	11 Q9JYLA	OC: Rhabditidae; Peloridiae; Caenorhabditis;	
8	59	64.1	328	5 Q9U336	OX: NCBL-TAXID=6239;	
9	59	64.1	380	2 Q9FB88	RN: [1] SEQUENCE FROM N.A. STRAIN=BRISTOL_N2; MEDLINE=99056613; PubMed=9851916;	
10	59	64.1	739	11 Q70575	RA: None; RT: "Genome sequence of the nematode <i>C. elegans</i> : a platform for investigating biology. The <i>C. elegans</i> Sequencing Consortium."; RL: Science 282:2012-2018(1998).	
11	59	64.1	751	11 Q9R1N9	RN: [2] RP: SEQUENCE FROM N.A. RC: STRAIN=BRISTOL_N2;	
12	59	64.1	1017	11 Q9X41	RA: Tin-Wollam A.; Fronick W.; RT: "The sequence of <i>C. elegans</i> cosmid C1BHT7."; RL: Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases. RN: [3]	
13	58	63.0	434	2 Q9ACN4	RP: SEQUENCE FROM N.A. RC: STRAIN=BRISTOL_N2;	
14	57	62.0	1051	5 Q26055	RA: Waterston R.; RT: "Direct Submission."; RL: Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases. DR: AF067607; AF098699; 1. -	
15	57	62.0	1352	13 Q9D7J0	DR: InterPro; IPR000087; InterPro; IPR02486; Col-cuticle_N. PFam: PF01391; Collagen_2.	
16	57	62.0	1419	11 Q63123	DR: Pfam: Col-cuticle_N; 1. KW: Hypothetical protein.	

SQ	SEQUENCE	460 AA;	41016 MW;	BF99CC80770E202 CRC64;	RA	de Wet W., Bernard M., Benson-Chanda V., Chu M., Dickson L., Weil D., Ramirez F.;
Query Match		75.0%;	Score 69;	DB 5;	Length 460;	"Organization of the human pro-alpha2(I) collagen gene.";
Best Local Similarity		80.0%;	Pred. No.	0.021;		J. Biol. Chem. 262:16032-16036(1987).
Matches	12;	Conservative	1;	Mismatches	2;	Indels 0;
Db	1	PAGPVGPNKGDKGVG 15				[3]
	237	PAGPPGPNGKDGAEAG 251				SEQUENCE OF 959-1351 FROM N.A.
RESULT 2						RX
ID	019470	PRELIMINARY;	PRT;	291 AA.	RA	MEDLINE=99304220; PubMed=2364107;
AC	Q19470;				RA	Makela J.K., Vuorio T., Vuorio E.;
DT	01-NOV-1996 (TREMBLrel. 01, last sequence update)				RT	"Growth-dependent modulation of type I collagen production and mRNA levels in cultured human skin fibroblasts.";
DT	01-DEC-2001 (TREMBLrel. 19, last annotation update)				RT	Biochim. Biophys. Acta 1049:171-176(1990).
DE	F15A2.1 PROTEIN.				RL	[4]
GN	F15A2.1.				RP	SEQUENCE FROM N.A.
OS	Caenorhabditis elegans.				RA	Dalgleish R.;
OC	Eukaryota; Metazoa; Chromadorea; Rhabditida; Rhabditoidea;				RL	Submitted (JUL-1996) to the EMBL/GenBank/DDBJ databases.
OC	Rhabditoidea; Peloderainae; Caenorhabditis.				RN	[5]
OX	NCBI_TaxID=6239;				RP	SEQUENCE FROM N.A.
RN	[1]	SEQUENCE FROM N.A.			RX	MEDLINE=97169339; PubMed=9016532;
RP	Gregory J.;				RA	Dalgleish R.;
RA	Submitted (MAR-1996) to the EMBL/GenBank/DDBJ databases.				RL	"The human type I collagen mutation database.";
RL	[2]	SEQUENCE FROM N.A.			DR	Nucleic Acids Res. 25:181-187(1997).
RP	MEDLINE=99069613; PubMed=9851916;				EMBL	274616; CA9A9869.1; -.
RA	none;				DR	InterPro; IPR00007; Collagen.
RT	*Genome sequence of the nematode <i>C.elegans</i> : A platform for investigating biology.";				DR	InterPro; IPR00885; Fib.collagen.C.
RT	Science 282:2012-2018(1998).				DR	Pfam; PF01410; COLFI; 1.
RL	EMBL; Z70207; CAA94128.1; -.				DR	Pfam; PF01391; Collagen; 18.
DR	InterPro; IPR002486; Collagen.				DR	ProDom; P002018; Fib.collagen_C; 1.
DR	Pfam; PF01391; Collagen; 2.				DR	SMART; SM00038; COLFI; 1.
DR	Pfam; PF01484; Col_cuticle_N.				KW	Signal; Collagen.
SO	SEQUENCE 291 AA; 28482 MW; BFG6AB31C7D5E2CA4 CRC64;				FT	Signal 1 22 POTENTIAL.
RX					SO	SEQUENCE 1366 AA; 129337 MW; 5796859E6E50286C CRC64;
RT					Q9AGC6	Query Match 68.5%; Score 63; DB 4; Length 1366;
RT					Q9AGC6	Best Local Similarity 73.3%; Pred. No. 0.52;
RT					AC	Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
RT					DT	01-JUN-2001 (TREMBLrel. 17, Created)
DR					DT	01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DR					DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DR					DE	COLLAGEN-LIKE PROTEIN B (FRAGMENT).
DR					OS	Streptococcus pyogenes.
SO					OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC					OC	Streptococcus.
OC					RN	NCBI_TaxID=1314;
RN					RP	[1]
RP					RP	SEQUENCE FROM N.A.
RC					STRAN=AP12;	
RA					RA	Rasmussen M., Björck L.;
RT					RT	"Phase-variation of a novel collagen-like surface protein in Streptococcus pyogenes."
RT					RL	Submitted (JAN-2001) to the EMBL/GenBank/DDBJ databases.
RL					DR	EMBL; AF336811; AAK0077.1; -.
DR					DR	InterPro; IPR00087; Collagen.
DR					PFAM	Pfam; PF01391; Collagen; 3.
FT					FT	NON_TER 1 1
FT					FT	NON_TER 326 326 AA; 33383 MW; BEES52296PF2445B CRC64;
SQ	SEQUENCE	326 AA;	33383 MW;	BEES52296PF2445B CRC64;	RA	Query Match 67.4%; Score 62; DB 2; Length 326;
RT					RT	Best Local Similarity 73.3%; Pred. No. 0.16;
RT					RT	Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
RL					RA	[2]
RL					RA	PARTIAL SEQUENCE FROM N.A.
RX					RA	MEDLINE=88058962; PubMed=2824475;





RP SEQUENCE FROM N.A.  
 RC TISSUE= MAMMARY TUMOR. WAP-TGF ALPHA MODEL. 7 MONTHS OLD. GROSS  
 RA TISSUE;  
 RA Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
 RL EMBL; B005481; AHA05481; 1; -.  
 DR InterPro; IPR000073; C1Q.  
 DR InterPro; IPR000087; Collagen.  
 DR Pfam; PF00386; C1Q; 1.  
 DR Pfam; PF01391; Collagen; 1.  
 DR SMART; SM00110; C1Q; 1.  
 SQ SEQUENCE 1017 AA; 107584 MW; 2EF903204DB9C80F CRC64;  
 Query Match 64.1%; Score 59; DB 11; Length 1017;  
 Best Local Similarity 73.3%; Pred. No. 1.5; Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
 Qy 1 PAGPMPNGKDGKV 15  
 Db 834 PAGPPGPGKDGQG 848

RESULT 13  
 Q9ACN4 PRELIMINARY; PRT; 434 AA.  
 ID Q9ACN4; 1  
 AC 09ACN4; 1  
 DT 01-JUN-2001 ('T-EMBLrel. 17, last sequence update)  
 DT 01-JUN-2001 ('T-EMBLrel. 17, last annotation update)  
 DE SCLB PROTEIN.  
 RN SEQUENCE FROM N.A.  
 OS Streptococcus Pyogenes.  
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
 OC Streptococcus.  
 OC NCBI\_TAXID=1314;  
 RN [1] SEQUENCE FROM N.A.  
 RP MEDLINE=21097281; PubMed=11158359;  
 RC STRAIN=PI80;  
 RA Whatmore A.M.; Pyoges sclb encodes a putative hypervariable surface  
 RT "Streptococcus Pyogenes sclb encodes a putative hypervariable surface  
 RT protein with a collagen-like repetitive structure.";  
 RL Microbiology 147:419-429 (2001).  
 DR EMBL; AJ301807; CAC3776; 1; -.  
 DR InterPro; IPR000087; Collagen.  
 DR InterPro; IPR001899; Gram\_Pos\_anchor.  
 DR Pfam; PF01391; Collagen; 4.  
 DR Pfam; PF00746; Gram\_pos\_anchor; 1.  
 DR PRINTS; PR00015; GPOSANCHOR.  
 DR PROSITE; PS00343; GRAM\_POS\_ANCHORING; UNKNOWN\_1.  
 KW Transmembrane.  
 SQ SEQUENCE 434 AA; 43885 MW; F77BB00DA881FD21 CRC64;

Query Match 63.0%; Score 58; DB 2; Length 434;  
 Best Local Similarity 73.3%; Pred. No. 0.86; Matches 11; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 PAGPMPNGKDGKV 15  
 Db 126 PAGPPGPGKDGQG 140

RESULT 14  
 Q9YJO0 PRELIMINARY; PRT; 1352 AA.  
 ID Q9YJO0; 1  
 AC 09YJO0; 1  
 DT 01-DEC-2001 ('T-EMBLrel. 19, last sequence update)  
 DT 01-DEC-2001 ('T-EMBLrel. 19, last annotation update)  
 DE PROCOLLAGEN TYPE I ALPHA 2 CHAIN.  
 RN [1] SEQUENCE FROM N.A.  
 OS Brachydanio rerio (Zebrafish) zebra danio.  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi;  
 OC Actinopterygii; Neopteri; Teleostei; Ostariophysi;  
 OC Cypriniformes; Cyprinidae; Danio.  
 OC NCBI\_TAXID=7955;  
 RN [1] SEQUENCE FROM N.A.  
 RA Morvan-Dubois G., Le Guellec D., Garrone R., Zylberberg L.,  
 RA Bonnard L.;  
 RT "Phylogenetic analysis on vertebrates fibrillar collagen enlightens  
 RT zebrafish a3(I) position and evidenced an evolutionary link between  
 RT collagen alpha chains and Hox clusters.";  
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
 RN [2] SEQUENCE FROM N.A.  
 RP MORVAN-DUBOIS G., HAFTEK Z., CROZET C., GARRONE R., LE GUELLEC D.,  
 RA "Structure and expression of the full length cDNA encoding zebrafish  
 RT alpha 2 type I collagen.";  
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AJ318213; CAC51030; 1; -.  
 KW Collagen.  
 SQ SEQUENCE 1352 AA; 12733 MW; 63D73C2CEAD44FE CRC64;

Query Match 62.0%; Score 57; DB 13; Length 1352;  
 Best Local Similarity 73.3%; Pred. No. 4; Matches 11; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 PAGPMPNGKDGKV 15  
 Db 1045 PAGPNPGKGDSNG 1059

RESULT 16  
 Q63123

ID	Q93123	PRELIMINARY;	PRT;	1419 AA.
AC	Q63123;	Q63156;		
DT	01-NOV-1996	(TREMBlrel. 01; Created)	DR	
DT	01-NOV-1996	(TREMBlrel. 01; Last sequence update)	DR	
DT	01-JUN-2001	(TREMBlrel. 17; Last annotation update)	DR	
DE	COLLAGEN ALPHA 1 TYPE II (T1 mRNA).			
T1	Rattus norvegicus (Rat).			
RA	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
RN	NCBI_TAXID=10116;			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=BONE FRACTURE CALLUS;			
RA	Urabe K.; Sarker G.; Bolander M.E.;			
RL	Submitted (Oct-1995) to the EMBL/GenBank/DBJ databases.			
RN	[2]			
RP	SEQUENCE OF 1372-1419 FROM N.A.			
RA	Wurtz T.; Brandstet C.; Lundmark C.; Christersson C.; Michaelson E.; Malmstrom V.; Reis S.; Engstrom A.; Burkhardt H.; Holmdahl R.;			
RT	"T cell recognition of carbohydrates on type II collagen.";			
J	J. Exp. Med. 180:745-749(1994).			
DR	EMBL; L48440; AAA97801; -.			
DR	EMBL; AU224879; CAA12179; 1; -.			
DR	X79816; CAA56213; 1; -.			
DR	InterPro; IPR000887; Collagen.			
DR	InterPro; IPR000885; fib_collagen_C.			
DR	Pfam; PRO1410; COLFI; 1.			
DR	ProDom; PD002078; fib_collagen_C; 1.			
DR	SMART; SM00214; VWC; 1.			
DR	PROSITE; PS01208; VWFC; 1.			
KW	Collagen.			
SQ	SEQUENCE OF 1419 AA; 134570 MW; B7C63B77819CE50B CRC64;			
Query Match	Best Local Similarity 73.3%; Score 57; DB 11; Length 1442; Matches 11; Conservative 0; Mismatches 4; Indels 0; Gaps 0;			
QY	1 PAGPWGPNGKDGKVG 15			
Db	871 PAGPPGPAGKDGPKG 885			
RESULT	18			
ID	Q62033	PRELIMINARY;	PRT;	1442 AA.
AC	Q62033;			
DT	01-NOV-1995	(TREMBlrel. 01; Created)	DR	
DT	01-NOV-1996	(TREMBlrel. 01; Last sequence update)	DR	
DT	01-DEC-2001	(TREMBlrel. 19; Last annotation update)	DE	
DE	PRO-ALPHA 1 TYPE II COLLAGEN.			
GN	COL2A1.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TAXID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=C57/BLACK;			
RA	Vuorio E.;			
RL	Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; M65161; AAA81021; -.			
DR	MGD; MGI:88452; COL2A1.			
DR	InterPro; IPR00087; Collagen.			
DR	InterPro; IPR00085; fib_collagen_C.			
DR	InterPro; IPR01007; VWFC.			
DR	Pfam; PRO1410; COLFI; 1.			
DR	Pfam; PRO1391; Collagen; 17.			
DR	ProDom; PD002078; fib_collagen_C; 1.			
DR	SMART; SM00214; VWC; 1.			
DR	PROSITE; PS01208; VWFC; 1.			
KW	SEQUENCE OF 1442 AA; 137829 MW; F0E7C11BCAF93B CRC64;			
Query Match	Best local Similarity 73.3%; Score 57; DB 11; Length 1442; Matches 11; Conservative 0; Mismatches 4; Indels 0; Gaps 0;			
QY	1 PAGPWGPNGKDGKVG 15			
Db	871 PAGPPGPAGKDGPKG 885			

RESULT 19

ID 062032 PRELIMINARY; PRT; 1459 AA.

AC 062032: 01-NOV-1996 (TREMBrel. 01, Created)  
01-NOV-1996 (TREMBrel. 01, Last sequence update)

DT 01-DEC-2001 (TREMBrel. 19, Last annotation update)

DE PRO-ALPHA-I TYPE II COLLAGEN.

GN COL2A1.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. [1]

RN SEQUENCE FROM N.A.

RP STRAIN=257/BLACK;

RX MEDLINE=91350489; PubMed=1805613;

RA Mesaranta M., Toman D., de Crombrugge B., Vuorio E.;

RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.

DR EMBL; M5161; AAAB0101.1; -.

DR MGD: MGI:88432; Col2a1.

DR InterPro; IPR000077; Collagen.

DR InterPro; IPR00085; Fib collagen\_C.

DR Pfam; PF01410; COLFI; 1.

DR Pfam; PF01391; Collagen; 18.

DR Pfam; PF0093; vwc; 1.

DR PRODOM; PDO0078; Fib collagen\_C; 1.

DR SMART; SM00038; COLIFI; 1.

DR SMART; SM0214; vwc; 1.

DR PROSITE; PS01208; WFBC; 1.

KW Collagen.

SQ SEQUENCE FROM N.A.

Query Match 62.0%; Score 57; DB 11; Length 1459;  
Best Local Similarity 73.3%; Pred. No. 4.4;  
Matches 11; Conservative 0; Mismatches 4; Indels 0; Gaps 0; Oy 1 PAGPWGPNGKDGKVG 15  
Db 888 PAGPPPGPAGKDGPKG 902

RESULT 20

ID 093208 PRELIMINARY; PRT; 303 AA.

AC 093208; 01-FEB-1997 (TREMBrel. 02, Created)  
01-DEC-2001 (TREMBrel. 19, Last annotation update)

DT 01-DEC-2001 (TREMBrel. 19, Last annotation update)

DE C14I1.1 PROTEIN.

C15I11.1.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Chromadorea; Rhabditida; Rhabditoidea; Rhaditidae; Pelodrinae; Caenorhabditis.

OX NCBI\_TAXID=6239; [1]

RP SEQUENCE FROM N.A.

RA Gardner A.E.; RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases. [2]

RP SEQUENCE FROM N.A. RP MEDLINE=99069013; PubMed=9831916; RA none;

RT "Genome sequence of the nematode *C.elegans*: A platform for investigating biology." Science 282:2012-2018 (1998).

RT DE InterPro; IPR00087; Collagen.

DR Pfam; PRO1391; Collagen. DR

SQ SEQUENCE 303 AA: 27889 MW; 9024926E86E07013 CRC64;

Query Match 60.9%; Score 56; DB 5; Length 303; Best Local Similarity 66.7%; Pred. No. 1.2; Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0; Oy 1 PAGPWGPNGKDGKVG 15  
Db 210 PAGPPGPAGKDGPKG 224

RESULT 21

ID 0905K6 PRELIMINARY; PRT; 588 AA.

AC 0905K6; 01-MAY-2000 (TREMBrel. 13, Created)  
01-MAY-2000 (TREMBrel. 13, Last sequence update)

DT 01-DEC-2001 (TREMBrel. 19, Last annotation update)

DE LATENT MEMBRANE PROTEIN 1 (FRAGMENT).

GN LMP1.

OS Cynomolgus Epstein-Barr Virus tsB-B6.

OC Gamma herpesvirinae; Lymphocryptovirus.

OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;

OX NCBI\_TAXID=104240; RN

RP SEQUENCE FROM N.A.

RX MEDLINE=90250747; PubMed=2160013;

RA Fujimoto K., Terato K., Miyamoto J., Ishiko H., Fujisaki M., Cho F., Honjo S.;

RT "Establishment of a B-lymphoblastoid cell line infected with Epstein-Barr-related virus from a cynomolgus monkey (*Macaca fascicularis*)."; J. Med. Primatol. 19:21-30(1990). [2]

RP SEQUENCE FROM N.A.

RA Faucher S., Wright K.E.; Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF181716; ADD56946.1; -.

DR InterPro; IPR002932; EGGSHELL.

DR InterPro; IPR000188; GABA<sub>A</sub> receptor.

DR Pfam; PRO2932; NEUR\_CHAN\_Memb; 1.

DR PRINTS; PRO1228; EGGSHELL.

FT NON\_TER 588  
SQ SEQUENCE 588 AA; 59338 MW; 62D85879901749BD CRC64;

Query Match 60.9%; Score 56; DB 12; Length 588; Best Local Similarity 60.0%; Pred. No. 2.4; Matches 9; Conservative 0; Mismatches 6; Indels 0; Gaps 0; Oy 1 PAGPWGPNGKDGKVG 15  
Db 470 PWGPWGPRCPSSNVG 484

RESULT 22

ID 090796 PRELIMINARY; PRT; 888 AA.

AC 090796; 01-NOV-1996 (TREMBrel. 01, Created)  
01-NOV-1996 (TREMBrel. 01, Last sequence update)

DT 01-DEC-2001 (TREMBrel. 19, Last annotation update)

DE ALPHA-1 TYPE XI COLLAGEN (FRAGMENT).

OS Gallus gallus (Chicken).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.

OC NCBI\_TAXID=9031; [1]

RP	SEQUENCE FROM N.A.						OX	NCBI_TAXID=5855;	
RC	STRAIN=WHITE LEISHORN; TISSUE=LIMB;						RN	[1]-	
RX	MEDLINE=93054557; PubMed=1429507;						RP	SEQUENCE FROM N.A.	
RA	Nah H.-D.; Barenbaum M.; Upholt W.B.;						RC	STRAIN=SALVADOR I;	
RT	"The chicken alpha 1 (XI) collagen gene is widely expressed in embryonic tissues.";						RX	MEDLINE=9608670; PubMed=8813677;	
RT	J. Biol. Chem. 267:22581-22586(1992).						RA	Kaslow D.C., Barr P.J., Crawford K.A., Holley L.J., Landsberg K.E., Gibson H.L., Kiefer M.C.,	
RL	EMBL; M8853; AAA48707.1; -.						RA	Barr P.J., Crawford K.A., Holley L.J., Landsberg K.E., Gibson H.L., Kiefer M.C.,	
DR	IntervPro; IPR000885; Fib_collagen_C.						RT	Identification and cloning of a locus of serine repeat antigen (sera)-related genes from Plasmodium vivax.;	
DR	Pfam; PF01410; COLEFI_1;						RL	Mol. Biochem. Parasitol. 78:55-65(1996).	
DR	Pfam; PF01391; Collagen_10.						DR	EMBL; U51723; AAB41487.1; -.	
DR	ProDom; PD002078; Fib_collagen_C; 1.						DR	IntervPro; IPR001064; Crystallin.	
DR	SMART; SM00038; COLFI_1.						DR	IPR000668; Peptidase_C1.	
FT	NON_TER	1	1				DR	Pfam; PF00112; Peptidase_C1; 1.	
FT	SEQUENCE	888 AA;	86409 MW;	B2EC9A0B20F903E	CRC64;		DR	PRINTS; PRO0705; PARIN.	
Query Match	Score 60.9%; Pred. No. 3.7%; Length 888;			Best Local Similarity 66.7%; Pred. No. 3.7%;	Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0;		DR	PROSITE; PS00235; CRYSTALLIN_BETAGAMMA; UNKNOWN_1.	
Qy	1 PAGPWPNKGDKGVG 15			Qy	1 PAGPWPNKGDKGVG 15		SQ	SEQUENCE 1076 AA; 115597 MW; 0B8C2509C3CC17AB CRC64;	
Db	308 PRGPOCPNGADGPQG 322			Query Match	Score 60.9%; Pred. No. 4.5%; Length 1076;		Query Match	Score 60.9%; Pred. No. 4.5%; Length 1076;	
Result 23				Best Local Similarity 66.7%; Pred. No. 4.5%;	Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;		Best Local Similarity 66.7%; Pred. No. 4.5%;	Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;	
044367				Qy	1 PAGPWPNKGDKGVG 15		Qy	1 PAGPWPNKGDKGVG 15	
ID	044367			Db	843 PDGPAGPSGDGNVG 857		Db	843 PDGPAGPSGDGNVG 857	
AC	044367;	PRELIMINARY;	PRT;	RESULT 23	09W7R9	PRELIMINARY;	RESULT 25	09W7R9	PRELIMINARY;
DT	01-JUN-1998 (TREMBLrel. 06, Last sequence update)			ID	09W7R9	PRELIMINARY;	ID	09W7R9	PRELIMINARY;
DT	01-JUN-1998 (TREMBLrel. 06, Last annotation update)			AC	09W7R9;		AC	09W7R9;	
DE	PRECOLLAGEN D.			DT	01-NOV-1999 (TREMBLrel. 12, Created)		DT	01-NOV-1999 (TREMBLrel. 12, Last sequence update)	
GN	PRECOL-D.			DT	01-NOV-1999 (TREMBLrel. 12, Last annotation update)		DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)	
OS	Mytilus edulis (Blue mussel); Bivalvia; Pteriomorphia; Mytilioida; Eukaryota; Metazoa; Mollusca; Mytilidae; Mytilus.			DE	ALPHA TYPE II COLLAGEN.		DE	ALPHA TYPE II COLLAGEN.	
OC				OS	Cynops pyrrhogaster (Japanese common newt).		OS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Caudata; Salamandroidea; Salamandridae; Cynops.	
OC				OC			OC		
OX	NCBI_TAXID=6550;			OX			OX		
RN	[1]-			RN	[1]-		RN	[1]-	
RP	SEQUENCE FROM N.A.			RP	SEQUENCE FROM N.A.		RP	SEQUENCE FROM N.A.	
RC	TISSUE=FOOT;			RC	TISSUE=FOOT;		RC	TISSUE=FOOT;	
RX	MEDLINE=9807044; PubMed=9405478;			RX	MEDLINE=9807044; PubMed=10474166;		RX	MEDLINE=9807044; PubMed=10474166;	
RA	Olin X.X., Coyne K.J., Waite J.H.;			RA	Asahina K., Obara M., Yoshizato K.;		RA	Asahina K., Obara M., Yoshizato K.;	
RT	"Tough tendons. Mussel byssus has collagen with silk-like domains.";			RT	"Expression of genes of type I and type II collagen in the formation and development of the blastema of regenerating newt limb.;"		RT	"Expression of genes of type I and type II collagen in the formation and development of the blastema of regenerating newt limb.;"	
RL	J. Biol. Chem. 272:32623-32627(1997).			RL	Dev. Dyn. 216:559-71(1999).		RL	Dev. Dyn. 216:559-71(1999).	
DR	EMBL; AB029249; AAC86638.1; -.			DR	EMBL; AB020465; BAB82043.1; -.		DR	EMBL; AB020465; BAB82043.1; -.	
DR	InterPro; IPR00087; Collagen.			DR	InterPro; IPR00087; Collagen.		DR	InterPro; IPR00087; Collagen.	
DR	pfam; PF01391; Collagen; 7.			DR	pfam; PF01410; COLFI_1.		DR	pfam; PF01410; COLFI_1.	
DR	ProDom; PD002078; Fib_collagen_C.			DR	ProDom; PD002078; Fib_collagen_C.		DR	ProDom; PD002078; Fib_collagen_C.	
KW	SEQUENCE. 922 AA; 80306 MW;	B2EC9A0B20F903E	CRC64;	KW	SEQUENCE. 1418 AA; 135067 MW;	C19A6E601A2K71E	KW	SEQUENCE. 1418 AA; 135067 MW;	C19A6E601A2K71E
Query Match	Score 60.9%; Pred. No. 3.8%; Length 922;			Query Match	Score 60.9%; Pred. No. 6; Length 1418;		Query Match	Score 60.9%; Pred. No. 6; Length 1418;	
Qy	1 PAGPWPNKGDKGVG 15			Best Local Similarity 66.7%; Pred. No. 6;	Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;		Best Local Similarity 66.7%; Pred. No. 6;	Matches 10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;	
Db	444 PAGPWPNKGDKGVG 458			Qy	1 PAGPWPNKGDKGVG 15		Qy	1 PAGPWPNKGDKGVG 15	
RESULT 24				Db	1096 PPGPVGPGSGKGDSNG 1110		Db	1096 PPGPVGPGSGKGDSNG 1110	
Q26154				RESULT 26	Q90W37	PRELIMINARY;	RESULT 26	Q90W37	PRELIMINARY;
ID	026154	PRELIMINARY;	PRT;	ID	Q90W37	PRELIMINARY;	ID	Q90W37	PRELIMINARY;
AC	026154;			AC	Q90W37;		AC	Q90W37;	
DT	01-NOV-1996 (TREMBLrel. 01, Created)			DT	01-DEC-2001 (TREMBLrel. 19, Created)		DT	01-DEC-2001 (TREMBLrel. 19, Last sequence update)	
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)			DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)		DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)	
DE	V-SERA 3.			DE	ALPHA 1 TYPE IIIA COLLAGEN PRECURSOR.		DE	ALPHA 1 TYPE IIIA COLLAGEN PRECURSOR.	
GN				GN			GN		
OS	Plasmodium vivax.			OS			OS		
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.			OC			OC		

GN COL2A1.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauvia; Aves; Neognathae; Galliformes; Phasianidae;  
 OC Gallus.  
 OX NCBI\_TAXID=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE="TETRALYMPHOCYTE"  
 RA Caixa X., Yongzhi X., Siqi G., Yiyang S.;  
 RL Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL: AV046949; AAC098621.1; -.  
 KW SIGNAL; Collagen.  
 FT SEQUENCE 1420 AA: 134999 MW: 88D9AB17F214FFF CRC64;  
 SQ Query Match Best Local Similarity 60.9%; Score 56; DB 13; Length 1420;  
 Matches 10; Conservative 66.7%; Pred. No. 6; Mismatches 1; Indels 4; Gaps 0;  
 DR QY 1 PAGPWNPGNGKDKGVG 15  
 Db 1098 PPGPVGPSSKGDSNG 1112  
 RN  
 RESULT 27  
 ID 091717 PRELIMINARY; PRT: 1486 AA.  
 AC DT 091717  
 AC DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 AC DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 AC DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 AC DE ALPHA-1 TYPE II COLLAGEN.  
 GN COL2A1.  
 OS xenopus laevis (African clawed frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi;  
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;  
 OC Xenopidae; Xenopus.  
 OX NCBI\_TAXID=8355;  
 RN  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92011898; PubMed=1918153;  
 RA Su M.W., Suzuki H.R., Bleker J.J., Solursh M., Ramirez F.;  
 RT "Expression of two nonallelic type II procollagen genes during Xenopus  
 laevis embryogenesis is characterized by stage-specific production of  
 alternatively spliced transcripts.";  
 RL J. Cell Biol. 115:565-575(1991).  
 DR EMBL: M63596; AAA49679.1; -.  
 DR SMART: SM00038; COLFI; 1.  
 DR PROSITE: PS01208; WFC; 1.  
 DR KW Collagen.  
 DR SO SEQUENCE 1491 AA: 142495 MW: 43026FFF08FB0314 CRC64;  
 DR Query Match Best Local Similarity 60.9%; Score 56; DB 13; Length 1491;  
 DR Matches 10; Conservative 66.7%; Pred. No. 6; Mismatches 1; Indels 4; Gaps 0;  
 DR QY 1 PAGPWNPGNGKDKGVG 15  
 Db 1170 PPGPVGPSSKGDSNG 1184  
 RN  
 RESULT 29  
 ID Q9UQT4 PRELIMINARY; PRT: 1767 AA.  
 AC DT Q9UQT4;  
 AC DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 AC DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 AC DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 AC DE COLLAGEN TYPE XI ALPHA-1.  
 GN COL11A1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TAXID=9606;  
 RN  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=04555728; PubMed=10486316;  
 RA Annunen S., Korppo J., Czarny M., Warman M.L., Brunner H.G.,  
 RA Kaarainen H., Mulliken J.B., Tranbjerg L., Brooks D.G., Cox G.F.,  
 RA Crusberg J.R., Curtis M.A., Davenport S.L.H., Friedrich C.A.,  
 RA Mokail S., Kaitila I., Kawczynski M.R., Larsen-Bielenski A.,  
 RA Olsen B.R., Shanno N., Somer M., Viikula M., Zlotogora J.,  
 RA Prockop D.J., Ala-Kokko L.;  
 RT "Splicing Mutations of 54 bp Exons in the COL11A1 Gene Cause Marshall  
 Syndrome but Other Mutations Cause Overlapping Marshall/Stickler  
 RT Phenotypes";  
 RT AM. J. Hum. Genet. 65:974-983(1999).  
 DR EMBL: AF010112; AAC04724.1; -.  
 DR EMBL: AF010179; AAC04724.1; JOINED.  
 DR EMBL: AF101080; AAC04724.1; JOINED.  
 DR EMBL: AF101081; AAC04724.1; JOINED.  
 DR EMBL: AF101082; AAC04724.1; JOINED.  
 DR EMBL: AF101083; AAC04724.1; JOINED.

AC 091718;  
 DT 01-Nov-1996 (TREMBLrel. 01, Created)  
 DT 01-Nov-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE ALPHA-1 TYPE II, COLLAGEN.  
 OS Xenopus laevis (African clawed frog).  
 OC Eukaryota; Metazoa; Craniata; Vertebrata; Buteleostomi;  
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;  
 OC Xenopidae; Xenopus.  
 OX NCBI\_TAXID=8355;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92011898; PubMed=1918153;  
 RA Su M.W., Suzuki H.R., Bleker J.J., Solursh M., Ramirez F.;  
 RT "Expression of two nonallelic type II procollagen genes during Xenopus  
 laevis embryogenesis is characterized by stage-specific production of  
 alternatively spliced transcripts.";  
 RT RL J. Cell Biol. 115:565-575(1991).  
 DR EMBL: M63596; AAA49679.1; -.  
 DR SMART: SM00038; COLFI; 1.  
 DR PROSITE: PS01208; WFC; 1.  
 DR InterPro: IPR000885; Fib\_Collagen\_C.  
 DR InterPro: IPR01007; WFC.  
 DR Pfam: PF01391; COLFI; 1.  
 DR Pfam: PF00093; WFC; 1.  
 DR PRODOM: PD002078; Fib\_Collagen\_C; 1.  
 DR SMART: SM00314; WFC; 1.  
 DR PROSITE: PS01208; WFC; 1.  
 DR KW Collagen.  
 DR SO SEQUENCE 1491 AA: 142495 MW: 43026FFF08FB0314 CRC64;  
 DR Query Match Best Local Similarity 60.9%; Score 56; DB 13; Length 1491;  
 DR Matches 10; Conservative 66.7%; Pred. No. 6; Mismatches 1; Indels 4; Gaps 0;  
 DR QY 1 PAGPWNPGNGKDKGVG 15  
 Db 1170 PPGPVGPSSKGDSNG 1184

